Journal of Digital Convergence
Vol. 17. No. 9, pp. 71-78, 2019

ISSN 1738-1916
https://doi.org/10.14400/JDC.2019.17.9.071

Exploring the Performance of Synthetic Minority Over-sampling Technique (SMOTE) to Predict Good Borrowers in P2P Lending

Francis Joseph Costello[1], Kun Chang Lee[2*]
[1]PhD Student, SKK Business School, Sungkyunkwan University
[2]Professor, Global Business Administration/Dept of Health Sciences & Technology, SAIHST Sungkyunkwan University

P2P 대부 우수 대출자 예측을 위한 합성 소수집단 오버샘플링 기법 성과에 관한 탐색적 연구

프란시스 조셉 코스텔로[1], 이건창[2*]
[1]성균관대학교 경영대학 박사과정, [2]성균관대학교 글로벌경영학과/삼성융합의과학원 융합의과학과 교수

Abstract This study aims to identify good borrowers within the context of P2P lending. P2P lending is a growing platform that allows individuals to lend and borrow money from each other. Inherent in any loans is credit risk of borrowers and needs to be considered before any lending. Specifically in the context of P2P lending, traditional models fall short and thus this study aimed to rectify this as well as explore the problem of class imbalances seen within credit risk data sets. This study implemented an over-sampling technique known as Synthetic Minority Over-sampling Technique (SMOTE). To test our approach, we implemented five benchmarking classifiers such as support vector machines, logistic regression, k-nearest neighbor, random forest, and deep neural network. The data sample used was retrieved from the publicly available LendingClub dataset. The proposed SMOTE revealed significantly improved results in comparison with the benchmarking classifiers. These results should help actors engaged within P2P lending to make better informed decisions when selecting potential borrowers eliminating the higher risks present in P2P lending.

Key Words : P2P lending, Predicting borrowers, Credit risk Assessment, Big Data, Classification, Synthetic Minority Over-sampling Technique

요 약 본 연구는 P2P 대부 플랫폼에서 우수 대출자를 예측시 유용한 합성 소수집단 오버샘플링 기법을 제안하고 그 성과를 실증적으로 검증하고자 한다. P2P 대부 관련 우수 대출자를 추정할 때 일어나는 문제점중의 하나는 클래스간 불균형이 심하여 이를 해결하지 않고서는 우수 대출자 예측이 쉽지 않다는 점이다. 이러한 문제를 해결하기 위하여 본 연구에서는 SMOTE, 즉 합성 소수집단 오버샘플링 기법을 제안하고 LendingClub 데이터셋에 적용하여 성과를 검증하였다. 검증결과 SMOTE 방법은 서포트 벡터머신, k-최근접이웃, 로지스틱 회귀, 랜덤 포레스트, 그리고 딥 뉴럴 네트워크 분류기와 비교하여 통계적으로 우수한 성과를 보였다.

주제어 : P2P 대부, 대출자 예측 신용위험평가, 빅데이터, 분류, 합성 소수집단 오버샘플링 기법

*Corresponding Author : Kun Chang Lee(kunchanglee@gmail.com)
Received June 19 2019 Revised July 20, 2019
Accepted September 20, 2019 Published September 28, 2019

1. Introduction

Social lending or otherwise known as Peer-to-peer (P2P) lending is slowly growing into a sizable portion of the lending environment thanks to the widespread access of the internet [1-3]. Companies such as Prosper, LendingClub (LC) and Kiva have allowed individuals to directly lend money to other individuals when small amounts is necessary for the borrower [4]. Within this lending environment, prospective lenders can fund listings made by prospective borrowers who have to specify the loan amount in order for prospective lenders to fulfill these listings [5]. In P2P lending, loans are usually uncollateralized meaning that lenders usually seek higher returns for their lent money as a compensation for the financial risk they incur by using these services [6]. Compared with traditional methods of lending, like in banks, P2P lending allows total autonomy to all actors and thus decisions are not influenced by any other actor. This makes P2P unique in its approach to lending.

When it comes to assessing the risks of P2P lending, an investor can look to traditional economic models as well as machine learning models [7-8]. Social lenders who interact with traditional financial credit score models usually do so while relying upon risk assessment models: usually externally produced by financial agencies [3]. Despite this, prior research has suggested that P2P lending works under different dynamics when compared to more traditional methods [4] and thus these risk assessment models may not yield desirable risk assessments [3].

The reasons for why the dynamics differ are as follows: (1) P2P lending platforms provide considerable amounts of data on borrowers: as little is known of borrowers on these P2P platforms, lenders can suffer from information asymmetry. This is where there is a lack of knowledge on the borrower. Companies such as LC provide as much data as possible on borrowers to promote otherwise blind loaning to unknown people. P2P lending is considerably more transparent compared to more traditional lending institutions [2]. (2) The loaning procedure can be imagined as an auction like process. A borrower puts up a listing and then lenders bid for the loan. The loans that have many bidders attract the most attention and can distract from other viable loan options on the website [3]. (3) Previous studies on P2P loaning have shown that the typical types of metrics used in traditional loan assessment are not always the best indicators.

When it comes to credit risk assessment, data sets inherently have unbalanced classes, with the minority class normally representing defaulted or written off loans [9]. However, up till now previous research studying P2P loans has not looked for methods in which to tackle this problem.

To identify good borrowers within the context of P2P lending this paper looks to tackle the problem of class imbalances through the implementation of an over-sampling technique known as Synthetic Minority Over-sampling Technique (SMOTE) [10]. To test our approach, we have implemented three classifiers (Support Vector Machines (SVM), Logistic Regression (LR) and K-Nearest Neighbor (kNN)) one ensemble technique (Random Forest (RF)) and one deep learning model (Deep Neural Network (DNN)). By implementing our approach to credit risk assessment, actors engaged in P2P lending should be able to make a more informed decision when choosing potential lenders leading to a reduction in risk. Furthermore, our research should help future researchers engaged in P2P lending credit risk assessment to look to sampling methodologies to help increase performance of risk models through big data and machine learning.

The rest of this paper is organized as follows: (1) a look at recent trends in P2P lending including common machine learning techniques, (2) a description of our methodology will be presented, (3) experimental results will be presented, and (4) lastly conclusion as well as future recommendation and limitations will be explored.

2. Related Work

When selecting an applicant for a loan, the risk of choosing a bad borrower outweighs the risk of selecting a bad borrower who turns out to be good [3]. Therefore, much effort in the literature has gone into exploring the development, application and evaluation of predictive decision support systems [8].

2.1 Machine learning and credit risk assessment

Early studies on credit scoring risk assessment, researchers mostly focused on single classifier problems implementing algorithms such as Neural Networks [11], SVM [12-14] and kNN [15]. Later, attention turned to the use of ensemble techniques with better results, as seen with RF (3), Decision Tree (DT)-based Bagging, Random Subspace [16] and Stacking [17]. More recently, deep learning has been introduced into credit risk assessment with Luo, Wu and Wu showing the effectiveness of implementing a Restricted Boltzmann machine-based deep belief network. This study was novel in its approach at assessing corporate credit rating risk and showed improved results compared to SVM and NN [18].

2.2 Class imbalances and SMOTE

When dealing with class imbalances there are two main approaches to deal with this problem: over/under sampling [19]. Under sampling looks to reduce the number of instances in the majority class; whereas, over sampling looks to increase the instances in the minority class. In this paper we implement the latter through SMOTE [10]. SMOTE creates instances by using a kNN algorithm to produce instances from the minority class. Where $S_{\min}, S_{maj}, S_{syn}$, are the classes for minority, majority and synthetic. The process: (1) Determine kNN for a sample $x_i INS_{\min}$ and determine the value of S_{syn}, (2) From a random sample $x_i(t=1,2...,k)$ from k nearest neighbors from the sample $x_i INS_{\min}$. (3) Implement (1) to create synthetic sample n_i This can be represented in the following way:

$$x_{nj} = x_{ij} + gap^*(x_{tj} - x_{ij}) \quad (1)$$

Where gap can be a random number between 0 and 1 with:

$$i = 1,2,..., |S_{\min}|, t = 1,2,...,k, j-1,2,...,m \quad (2)$$

SMOTE generates as many synthetic instances in a dataset as in the minority class [20].

2.3 Classifying Models

Next, we will present the models in which we are implementing in our paper.

2.3.1 Support Vector Machines

Introduced by Boser et al. (1992) [21], SVM is a classification and regression model where there exists a hyper plane whereby the decision plane is separated by the positive (+1) and negative (-1) classes [22]. Given a training data set, $(x1, y1)$, $(x2, y2)...(x_m y_m)$ in which $x_i INR^d$ signifies vectors in a d-dimensional hyper plane, and $y_i \in [-1,+1]$ is a class label given to the data. SVM are then represented by morphing the input vectors into a new, higher dimensional analogue plane indicated as: $\Phi: R^d \rightarrow H^f$ in which $d < f$. Thereafter, an optimum hyperplane is formed by a kernel function $K(x_i, x_j)$, which is the product of the input vectors x_i and x_j, in which $K(x_i, x_j) = \Phi(x_i)^* \Phi(x_j)$ [23]. In this paper we implement the polynomial SVM, where P is the degree of polynomial:

$$K_{poly}(x_i, x_j) = (x_i, x_j + 1)^P \quad (3)$$

2.3.2 Logistic Regression

LR is popular and widely used model in credit risk assessment [9] and P2P lending [1]. The approach for LR can be seen for binary classification in the following formula:

$$\hat{f}_{LR}(X) := \frac{e^{\acute{\beta}X}}{1+e^{\acute{\beta}X}} \tag{4}$$

Where $\acute{\beta_i}$ is calculated though the maximum likelihood method [24].

2.3.3 K-Nearest Neighbor

kNN is a type of instance-based learning that implements a similarity function in order to classify training data [25] where k is set search for nearest neighbors. This k number searches instances in groups based on k and uses a simple majority vote [24].

$$simularity(x,y) = \sqrt{\sum_{i-1}^{n} f(x_i,y_i)} \tag{5}$$

Where $f(x_i,y_i) = (x_i - y_i)^2$ is for numerical values.

2.3.4 Random Forest

RF is a type of assembling classifier that builds randomized decision trees for learning. This follows a divide-and-conquer approach whereby the root node relates to the entire training data. In this paper, splitting of the nodes is performed based on the Gini index [3]:

$$G(X_i) := \sum_{j=1}^{j} \Pr(X_1 = L_j)(1 - \Pr(X_1 = L_j)) = 1 - \sum_{j=1}^{j} \Pr(X_1 = L_j)^2$$
(6)

2.3.5 Deep Neural Networks

DNNs are a type of NN that have come to prominence in academia of late. DNNs are based on feedforward networks which can be represented by composing together many different functions say f^1, f^2, f^3, are all interconnected within a given chain:

$$f(x) = f^{(3)}(f^{(2)}(f^1)(x))) \tag{7}$$

Here, $f^{(1)}$ represents the input layer, f^2 represents a deep/hidden layer, and f^3 represents an output layer. DNNs allow for a nonlinear transformation represented by Φ. Φ is a way of describing x based on a number of features within a given DNN, whereby learning Φ is the ultimate goal [26]. The model $y = f(x:\theta,w) = \Phi(x:\theta)$ can be seen as θ parameters that are implemented to learn Φ from a broad class of functions, and parameters w that go from $\Phi(x)$ to an output desired by the user. Φ can be viewed as the deep (hidden) layer of a given DNN meaning a user only needs to find the right general function rather than needing a precise definition of a given model. This makes DNN versatile and thus suitable for many areas of research [27].

3. Methodology

In this section a brief description of the methodology that we used in order to study P2P risk.

3.1 LendingClub dataset

The dataset in this study was retrieved from LC as it is publicly available. The data we obtained was based upon all loan requests in the year 2016. Following recommendations from previous researchers [3] we applied the same steps to create the final feature list. The final data set included 18 features and one class. All pre-processing and data manipulation followed recommendations from Malekipirbazari & Aksakalli (2015) [3]. Features selected included: (1) Loan status (class attribute), (2) annual income, (3) credit age, (4) delinquencies (last two years), (5) employment length, (6) home ownership, (7) inquiries (credit inquiries in the last six months), (8) loan amount, (9) loan purpose, (10) open accounts (currently opened credit lines), (11) total accounts (total number of credit line held), (12) term (length of loan).

The next features are ratios in the data or made previously in literature [3]: (13) DTI (debt to income ratio), (14) Income to Payment Ratio: this ratio represents the loan's monthly payments to monthly income, (15) Revolving Utilization Rate, (16) Revolving to Income Ratio: revolving credit balance to the borrower's monthly income.

The last two are based upon scores provided by LC. (17) FICO Score: This is a standard credit line that is used in the majority of lending decisions in the US [4]. It is based on financial attributes from the borrowers' credit records [3]. (18) LC Grade: this is grade given by the LC themselves. These are A1-G5, with A representing a less risky loan.

3.2 Pre-processing

In order to prepare the data for analysis we had to convert the nominal data into binary numerical figures. This meant that the data set was left with 12 original numerical features and a further 18 numerical features that resulted from the binarization. Furthermore, all data was standardized to allow for all features to have a zero mean and unit standard deviation.

3.3 Cost sensitive analysis

In credit risk assessment the aim is to try and find bad borrowers, and thus a classifier should hold great risk in misclassifying a bad borrower [3]. For this reason, we implemented the use of a cost-sensitive analysis with a RF algorithm to compare with [3].

3.4 Model description

All tests were performed using the WEKA too l[1]. Firstly, the baseline SVM, LR and kNN models from the WEKA tool were used. For RF, and based on recommendation by [3], the model implemented maximum tree depth of 25, maximum tree build of 80 and also a tree split number of 5. Lastly, the implemented DNN model was made with four dense layers that consisted of nodes made of 32, 64, 128 and 256 shaped deep layers, with a softmax output layer. The activation function used was the rectified linear unit (ReLu).

4. Results

In this paper, we implemented four common algorithms seen within P2P lending literature as well as implemented a deep learning NN. Our test was to find out whether our implementation of SMOTE can lead to greater results based on two main metrics: accuracy (ACC) and the area under the roc curve (AUC). All tests were done with the implementation of 10-fold cross validation whereby the training data was split into 10 subsets of equal size. Next we tested the SMOTE model against the original model [3] with t-tests.

4.1 Cost-sensitive analysis

Malekipirbazari & Aksakalli suggest within their methodology that the cost sensitive ratio should be 5:1 (2015) [3]. However, with the implementation of SMOTE, the minority class becomes more prominent in the dataset. Therefore, 5:1 seems too harsh.

Therefore, we tested cost sensitivity from 5:1 with the SMOTE dataset based on a RF classifier. When it comes to cost-sensitive analysis there is a trade-off between the accuracy and precision. As can be seen in these results with 3:1 a reduction in 2.46% precision allows for an accuracy increase of 6.59%. In our model will use 3:1, where misclassifying bad borrowers has three times more weight.

1) https://www.cs.waikato.ac.nz/ml/weka

Table 1. This table shows the evaluation of the Cost−sensitivity algorithm implemented with the Random Forest algorithm to find a good fit.

Cost−sensitivity Analysis with Random Forest			
	ACC*	AUC**	Precision
RF 3:1	79.25%	0.86	87.88%
RF 4:1	75.79%	0.85	89.32%
RF 5:1	72.66%	0.85	90.34%
* Accuracy, ** Area Under the Curve - AUC			

4.2 Empirical results

As can be seen in table 2, with the implementation of SMOTE and a cost−sensitivity of 3:1, our proposed method produces improved accuracy and greater AUC results. This means that our model now has much more confidence in achieving the classification of 79.25% compared with the originally proposed model [3]. Furthermore, coherent with previous P2P lending research RF outperforms all the other classification models, agreeing with [3].

Table 2. This table shows the results from implementing the five algorithms on the original dataset and the dataset which was over−sampled with SMOTE.

	Original Data		SMOTE	
	ACC	AUC	ACC	AUC
RF	77.87%	0.70	79.25%	0.86
SVM	59.96%	0.65	64.90%	0.65
LR	64.95%	0.71	64.95%	0.71
kNN−10	57.12%	0.62	62.14%	0.76
kNN−100	54.63%	0.70	60.85%	0.67
DNN	59.16%	0.71	61.36%	0.70

Table 3. This table shows the results from implementing T−tests based on the Original/SMOTE dataset with Random Forest as the classifier.

	Mean	Std. Dev	Std. Error Mean	t	Sig. 2−tailed
RF_Ori_acc - RF_SMO Accuracy***	−1.382	.445	.140	−9.808	.000
RF_Ori_auc - RF_SMO AUC***	−.1601	.006	.002	−76.88	.000

Furthermore, the implication of SMOTE has shown considerably greater results compared with the baseline models especially in the AUC metric. This is further seen in the t−test of RF whereby both ACC and AUC are significant in performance comparably.

5. Conclusion

In this paper we have analyzed recent trends in social lending (P2P) using a classification methodology. Specifically, we implemented the use of the Synthetic Minority Over−sampling Technique (SMOTE) [10] in order to help reduce the imbalanced class problem and found greater results in this model when applied alongside a Random Forest ensemble classifier [3]. With this, lenders engaged with P2P lending should now be better placed to identify good and bad borrowers and thus make a more informed decision. with this, lenders will reduce their risk and help to improve their portfolio.

Our research has introduced the idea of synthetically enhancing the minority data set within P2P credit risk analysis and thus should lead other researchers to also start to explore a similar methodology when it comes to researching P2P lending. Future work can look into exploring other sampling techniques like Random under−sampling or the One−Sided Selection technique.

Limitations of this research can be seen in the fact that we only used data from 2016. Results from other years could help to validate whether this methodology is valid or not. Another limitation is that this methodology may be limited to the LC data only and therefore this model should be evaluated on another P2P lending data sets.

REFERENCES

[1] H. Zhao, Y. Ge, Q. Liu, G. Wang, E. Chen, & Zhang.

(2017). P2P lending survey: platforms, recent advances and prospects. *ACM Transactions on Intelligent Systems and Technology (TIST), 8(6),* 72.
DOI : 10.1145/3078848

[2] C. Serrano-Cinca & B. Gutiurrez-Nieto. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems, 89,* 113-122.
DOI : 10.1016/j.dss.2016.06.014

[3] M. Malekipirbazari & V. Aksakalli. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications, 42(10),* 4621-4631.
DOI : 10.1016/j.eswa.2015.02.001

[4] R. Emekter, Y. Tu, B. Jirasakuldech & M. Lu. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics, 47(1),* 54-70.
DOI : 10.1080/00036846.2014.962222

[5] Y. Guo, W. Zhou, C. Luo, C. Liu & H. Xiong. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research, 249(2),* 417-426.
DOI : 10.1016/j.ejor.2015.05.050

[6] A. Byanjankar, M. Heikkill & J. Mezei. (2015, December). Predicting credit risk in peer-to-peer lending: A neural network approach. *In 2015 IEEE Symposium Series on Computational Intelligence* (pp. 719-725). IEEE.
DOI : doi.org/10.1109/SSCI.2015.109

[7] J. Abellion & J. G. Castellano. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications, 73,* 1-10.
DOI : 10.1016/j.eswa.2016.12.020

[8] S. Lessmann, B. Baesens, H. V. Seow & L. C. Thomas. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247(1),* 124-136.
DOI : 10.1016/j.ejor.2015.05.030

[9] I. Brown & C. Mues. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39(3),* 3446-3453.
DOI : 10.1016/j.eswa.2011.09.033

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall & W. P. Kegelmeyer. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research, 16,* 321-357.
DOI : 10.1613/jair.953

[11] K. K. Lai, L. Yu, S. Wang & L. Zhou. (2006, September). Credit risk analysis using a reliability-based neural network ensemble model. *In International Conference on Artificial Neural Networks* (pp. 682-690). Springer, Berlin, Heidelberg.
DOI : 10.1007/11840930_71

[12] L. Yu, X. Yao, S. Wang & K. K. Lai. (2011). Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. *Expert Systems with Applications, 38(12),* 15392-15399.
DOI : 10.1016/j.eswa.2011.06.023

[13] L. Yu, S. Wang & J. Cao. (2009). A modified least squares support vector machine classifier with application to credit risk analysis. *International Journal of Information Technology & Decision Making*, 8(04), 697-710.
DOI : 10.1142/S0219622009003600

[14] Y. Wang, S. Wang & K. K. Lai. (2005). A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*, 13(6), 820-831.
DOI : 10.1109/TFUZZ.2005.859320

[15] W. E. Henley & D. J. Hand. (1996). A K-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 45(1), 77-95.
DOI : 10.2307/2348414

[16] G. Wang, J. Ma, L. Huang & K. Xu. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, 26, 61-68. DOI : 10.1016/j.knosys.2011.06.020

[17] G. Wang, J. Hao, J. Ma & H. Jiang. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert systems with applications, 38(1),* 223-230.
DOI : 10.1016/j.eswa.2010.06.048

[18] C. Luo, D. Wu & D. Wu. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, 65, 465-470.
DOI : 10.1016/j.engappai.2016.12.002

[19] J. Sun, J. Lang, H. Fujita & H. Li. (2018). Imbalanced enterprise credit evaluation with DTE-SBD: decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. *Information Sciences*, 425, 76-91.
DOI : 10.1016/j.ins.2017.10.017

[20] L. Zhang & W. Wang. (2011, September). A re-sampling method for class imbalance learning with credit data. *In 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, 1,* 393-397.
DOI : 10.1109/ICM.2011.34

[21] B. E. Boser, I. M. Guyon & V. N. Vapnik. (1992, July). A training algorithm for optimal margin classifiers. *In Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). ACM.
DOI : 10.1145/130385.130401

[22] H. Byun & S. W. Lee. (2003). A survey on pattern recognition applications of support vector machines. *International Journal of Pattern Recognition and Artificial Intelligence, 17(3),* 459-486.

DOI : 10.1142/S0218001403002460

[23] M. Huang, C. Chen, W. Lin, S. Ke & C. Tsai. (2017). SVM and SVM Ensembles in Breast Cancer Prediction. *PLOS ONE*, 12(1), 1-14.
DOI : 10.1371/journal.pone.0161501

[24] I. H. Witten, E. Frank, M. A. Hall & C. J. Pal. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

[25] D. W. Aha, D. Kibler & M. K. Albert. (1991). Instance-based learning algorithms. *Machine learning, 6(1),* 37-66.
DOI : 10.1007/BF00153759

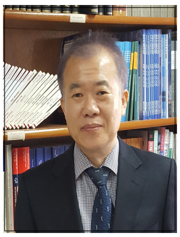
[26] Y. LeCun, Y. Bengio & G. Hinton. (2015). Deep learning. *Nature, 521(7553),* 436.
DOI : 10.1038/nature14539

[27] I. Goodfellow, Y. Bengio & A. Courville. (2016). *Deep learning*. MIT press.

프란시스 조셉 코스텔로(Francis J. Costello) [정회원]

· 2019년 2월 : 성균관대 경영대학 일반대학원 프런티어 경영 석사
· 2019년 2월 ~ 현재 : 성균관대학교 경영대학 일반대학원 박사과정
· 관심분야 : 데이터 마이닝, 감정분석, Agent-based Modeling 등
· E-Mail : joe.costehello@gmail.com

이 건 창(Kun Chang Lee) [정회원]



· 1984년 2월 : 카이스트 경영과학과(공학석사-의사결정지원)
· 1988년 8월 : 카이스트 경영과학과(공학박사-인공지능)
· 성균관대학교 경영대학 및 삼성융합의과학원 (SAIHST) 융합의과학과 교수
· 관심분야 : 창의성과학, 인공지능, 헬스 인포매틱스, 감성분석 등
· E-Mail : kunchanglee@gmail.com