2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)
4-7 March 2018
Las Vegas, Nevada, USA

Toward Personalized Sleep-Wake Prediction from Actigraphy

Aria Khademi[1], Yasser EL-Manzalawy[1], Orfeu M. Buxton[2,3,4,5], and Vasant Honavar[1]

Abstract— **Actigraphy offers a low-cost alternative to conventional polysomnography (PSG) for screening of sleep-wake patterns. Effective use of actigraphy signals requires reliable methods for detecting sleep-wake states from actigraphy measurements. Hence, there is a growing interest in machine learning methods for training predictive models of sleep-wake states from actigraphy data. Existing work has focused on training a single predictive model for the entire population. However, accounting for individual differences, such as age, biological factors, or lifestyle-related variations, calls for personalized models for reliable identification of sleep-wake states from actigraphy data. This study investigates whether personalized models, trained on individual data, can match the performance of generalized models trained on population data. Using a dataset of 54 individuals, we systematically trained and tested personalized and generalized sleep-wake detectors developed using five commonly used machine learning algorithms. Results of our experiments show that personalized sleep-wake predictors are competitive, in terms of their predictive performance, with their generalized counterparts. Our work demonstrates the feasibility of developing reliable personalized sleep-wake states predictors from actigraphy data. This study lays the groundwork for developing personalized models for sleep-wake states detection that are better equipped to handle individual differences.**

I. INTRODUCTION

Sleep impacts virtually all aspects of life, including health, mood, emotions, cognition, memory, behavior, and performance. Polysomnography (PSG), together with clinical assessments, is the *de facto* gold standard for assessing sleep quality and diagnosing sleep disorders [1]. However, the cost of PSG recordings and the need for expert scoring of PSG signals to identify sleep-week states can limit the applicability of PSG in large-scale population-level sleep studies, especially those aimed at characterizing individual differences in sleep patterns under real-world conditions.

Actigraphy, which measures movement using an accelerometer, often augmented with additional sensors, offers a relatively inexpensive, portable, convenient, and non-invasive alternative to PSG for screening of sleep disorders. Actigraphy's validity in assessment of sleep has been recognized by American Sleep Disorders Association and extensive research shows growing interest in using actigraphy for studies on sleep health [2]. However, effective use of actigraphy requires methods for reliable detection of sleep-wake states from actigraphy data. Hence, a variety of methods have been developed for this problem including sleep scoring algorithms [3]–[16], and sleep-wake states predictors trained on datasets of expert-annotated actigraphy data using machine learning methods [17]–[19].

We note that existing methods for sleep-wake states prediction rely on a single algorithm, or a single predictive model, which we refer to as a *generalized model*, for the entire population. However, several studies have reported considerable variability in sleep patterns across individuals due to characteristics such as age, lifestyle, biological factors, and prescription drug use [20], [21]. Accounting for such variability calls for *personalized* sleep-wake states predictors. Such personalized models offer the additional advantage of being amenable to being trained *in situ* on each individual's mobile device, thus eliminating the need for aggregating and analyzing the data in a central repository, and enhancing privacy and security.

Against this background, this study investigates whether personalized sleep-wake states predictors that are trained on individual data can match the performance of their generalized counterparts that are trained on population data. The results of our experiments show that personalized models can achieve performance that is competitive with their generalized counterparts. Furthermore, for approximately 30% of the individuals, personalized models had greater prediction accuracy than their generalized counterparts. Hence, these results establish the feasibility of developing reliable personalized sleep-wake states predictors from actigraphy data.

II. MATERIALS AND METHODS

A. Data

The data used in this study are obtained from [9] and consist of simultaneous actigraphy and PSG recordings obtained in the sleep laboratory, from healthy adults (baseline) [22]–[24], older adults (previously unpublished data described in [9]), sleep restriction in healthy participants [25] (sleep recordings were in controlled standardized conditions and study participants received either no medication or placebo on individual nights), and daytime sleep of night-workers [26]. In all cases, participants spent 8.5 hours in bed except in the sleep restriction study which included 10 hours in bed

This project was supported in part by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR000127 and TR002014, and the Center for Big Data Analytics and Discovery Informatics at Pennsylvania State University, the Edward Frymoyer Endowed Professorship in Information Sciences and Technology at Pennsylvania State University and the Sudha Murty Distinguished Visiting Chair in Neurocomputing and Data Science at the Indian Institute of Science [both held by Vasant Honavar]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

[1]College of Information Sciences and Technology, Pennsylvania State University, University Park, PA (email: khademi@psu.edu)

[2]Department of Biobehavioral Health, Pennsylvania State University, University Park, PA

[3]Division of Sleep Medicine, Harvard Medical School, Boston, MA

[4]Department of Social and Behavioral Sciences, Harvard Chan School of Public Health, Boston, MA

[5]Department of Medicine, Brigham and Womens Hospital, Boston, MA

978-1-5386-2405-0/18/$31.00 ©2018 IEEE

on each of the sleep-replete nights and 5 hours in bed on sleep restricted nights. For each participant, we have data from multiple sleeping periods, with each sleeping period consisting of roughly 1000 epochs of 30-second duration each. Each such epoch is assigned a sleep/wake label based on expert annotation of the PSG signal for the same epoch. Because our exploratory analyses showed that we need data from ≥ 2 sleeping periods for training the predictors and we need at least 1 additional sleeping period for testing the predictors, we limited our analyses to 54 participants with data for a minimum of 3 sleeping periods. To factor out any night-specific effects, we randomly partitioned data from the 54 participants into three disjoint groups, each with 18 participants. For each participant, we designated one of the first 3 sleeping periods for testing and all of the remaining sleeping periods for training the classifier(s), with the group membership determining which, among the first 3 nights, to set aside for testing. We normalized the raw actigraphy measurements to fall in the interval [0,1] for each sleeping period so as to minimize the effect of variability across devices and/or across participants.

B. Feature Extraction

Based on the results of an initial exploratory analysis, we used the features extracted from a sliding window of 21 actigraphy epochs, centered at the target epoch, as input to the sleep-wake states predictor. From each window, we extracted 18 features used in previous studies [16], [27]: 10^{th}, 20^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles, mean, sum of values, standard deviation, coefficient of variation, peak-to-peak amplitude, inter quartile range, skewness, kurtosis, signal power, peak intensity, zero crossings, time above threshold, and maximum value, along with the 21 normalized actigraphy measurements. Thus, each sliding window was encoded by a 39-tuple of feature values; and labeled as either *sleep*, or *wake*.

C. Predictive Models

We experimented with 5 commonly used machine learning algorithms all using Python's sklearn (version 0.17.1) implementations (with the default parameter settings unless stated otherwise): 1) Naive Bayes (NB) [28] with a multi–variate Bernouli model for the likelihood and default sklearn discretization scheme for continuous values, 2) Regularized logistic regression (RLR) [29] with an L2 regularizer and stochastic gradient descent (SGD) used to optimize the objective function, 3) Random forest (RF) [30] with 100 trees, 4) AdaBoost (AB) [31] with a 100 decision trees, and 5) Extreme gradient boosting (XGB) [32], [33] with 100 trees, trained using a logistic loss function, and an L2 regularization term.

D. Performance Evaluation

We trained personalized sleep-wake states predictors, one for each individual, using the individual's training data. We also trained generalized sleep-wake states predictors using the training data from the entire population. In both cases, to permit direct comparison, the resulting predictors were evaluated on the test data for each individual. Designating wake and sleep states as the positive and negative classes, respectively, we used the standard performance measures of accuracy (ACC), sensitivity (SN), specificity (SP), Matthews correlation coefficient (MCC) [34], and the area under the Receiver Operating Characteristic (ROC) curve [35] to evaluate the performance of the resulting binary (sleep/wake) classifiers.

III. Results

A. How do personalized sleep-wake predictors compare with their generalized counterparts?

For each choice of learning algorithm (See Section II. C), we compare the performance of the generalized sleep-wake predictors with their personalized counterparts in terms of ACC, SN, SP, MCC, and AUC. Table 1 shows performance estimates averaged over individuals drawn from all studies. We observe that the performance of the personalized sleep-wake states predictors, as measured by the average AUC estimated from test data, is comparable to that of their generalized counterparts, a conclusion that is borne out by comparison of the corresponding ROC plots (Fig. 1, darker orange and blue curves). We further observe from Table 1 that the XGB sleep-wake predictors outperform their NB, RLR, AB, and RF counterparts with an AUC of 0.84. Table 2 shows performance estimates averaged over individuals within each study. We observe that in each case, the performance of personalized models is competitive with that of their generalized counterparts. Furthermore, we found no significant correlations between the difference in the performance (AUC) of personalized models trained using XGB and their generalized counterparts, and the individual characteristics (age, gender, sleep disorder, and time in bed).

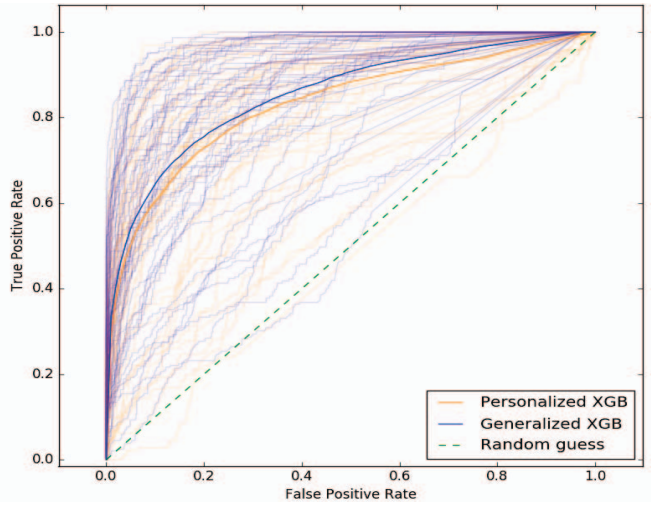

Fig. 1. ROC curves of Personalized and Generalized XGB sleep-wake states predictors. The two darker ROC curves show the performance of the generalized predictor (blue) and personalized predictors (orange) averaged over all individuals across all studies. The lighter curves show performance of the generalized XGB predictor and personalized XGB predictors tested on each individual.

TABLE I

PERFORMANCE OF PERSONALIZED AND GENERALIZED SLEEP-WAKE STATES PREDICTORS AVERAGED OVER ALL PARTICIPANTS ACROSS ALL STUDIES COMBINED.

	Personalized models					**Generalized models**				
Classifier	ACC	SN	SP	MCC	AUC	ACC	SN	SP	MCC	AUC
NB	0.75	**0.74**	0.74	0.38	0.83	0.75	**0.69**	0.77	0.36	0.83
RLR	**0.86**	0.29	**0.98**	0.36	0.82	0.86	0.26	**0.99**	0.35	0.83
RF	0.85	0.45	0.93	0.40	0.81	0.86	0.41	0.96	0.41	0.80
AB	**0.86**	0.46	0.95	0.44	0.80	**0.87**	0.37	0.98	**0.45**	**0.85**
XGB	**0.86**	0.45	0.95	**0.45**	**0.84**	**0.87**	0.38	0.98	**0.45**	**0.85**

TABLE II

PERFORMANCE OF PERSONALIZED AND GENERALIZED SLEEP-WAKE STATES PREDICTORS AVERAGED SEPARATELY ACROSS INDIVIDUALS WITHIN EACH STUDY: SLEEP RESTRICTION (SR), OLDER ADULTS ON MEDICATION (TI), ACOUSTICS DISTURBED (AC), AND NIGHT WORKERS (NW).

		Personalized models					**Generalized models**				
Study	Classifier	ACC	SN	SP	MCC	AUC	ACC	SN	SP	MCC	AUC
SR	NB	0.76	**0.75**	0.75	0.40	0.85	0.68	**0.79**	0.66	0.34	0.85
	RLR	**0.88**	0.32	**0.98**	0.42	0.86	0.88	0.33	**0.98**	0.41	0.86
	RF	0.87	0.47	0.94	0.44	0.85	0.87	0.50	0.95	0.47	0.85
	AB	**0.88**	0.49	0.95	**0.49**	0.85	0.88	0.46	0.97	0.50	**0.89**
	XGB	0.84	0.43	0.94	0.42	**0.88**	**0.89**	0.48	0.97	**0.51**	**0.89**
TI	NB	0.80	**0.67**	0.80	0.38	0.84	0.75	**0.74**	0.75	0.39	0.83
	RLR	0.86	0.23	**1.00**	0.35	0.85	0.85	0.20	**0.99**	0.27	0.84
	RF	0.89	0.52	0.95	0.49	0.84	0.87	0.40	0.96	0.41	0.83
	AB	**0.90**	0.48	0.98	**0.53**	0.86	0.87	0.37	0.98	0.45	0.87
	XGB	0.80	0.27	0.95	0.32	**0.88**	**0.88**	0.37	0.98	**0.46**	**0.88**
AC	NB	0.60	**0.70**	0.57	0.24	**0.70**	0.72	**0.41**	0.83	**0.25**	**0.69**
	RLR	0.76	0.13	**0.98**	0.18	0.67	**0.77**	0.10	**1.00**	0.23	0.68
	RF	0.72	0.26	0.88	0.17	0.64	0.76	0.17	0.96	0.21	0.63
	AB	0.75	0.26	0.92	0.22	0.62	**0.77**	0.15	0.98	**0.25**	**0.69**
	XGB	**0.87**	0.47	0.96	**0.48**	0.66	**0.77**	0.16	0.98	**0.25**	**0.69**
NW	NB	0.82	**0.79**	0.82	0.45	0.90	0.83	**0.75**	0.84	0.45	0.90
	RLR	**0.91**	0.40	**0.98**	0.44	0.89	0.91	0.33	**0.99**	0.42	0.90
	RF	**0.91**	0.53	0.95	0.49	0.88	0.91	0.49	0.96	0.48	0.86
	AB	**0.91**	0.55	0.95	0.50	0.84	**0.92**	0.44	**0.99**	**0.54**	**0.92**
	XGB	**0.91**	0.55	0.96	**0.52**	**0.91**	**0.92**	0.45	**0.99**	**0.54**	**0.92**

B. Do personalized sleep-wake predictors outperform their generalized counterparts in some individuals?

Interestingly, in the case of 16 out of the 54 participants, or roughly 30% of the population, personalized XGB significantly outperforms generalized XGB in terms of AUC ($p - value < 0.00014$ in t-test). Moreover, in the case of 8 other individuals, performance of personalized XGB matches that of generalized XGB (Data not shown). Thus, in approximately 44% of the individuals, personalized XGB outperforms or matches the performance of generalized XGB. This raises the possibility of optimally combining personalized and generalized models to further improve the reliability of sleep-wake states detection.

IV. SUMMARY AND CONCLUSIONS

Accounting for differences in sleep patterns across individuals calls for personalized sleep-wake states predictors. The results of this study have established the feasibility of developing such personalized models that are competitive with their generalized counterparts. These results open up the possibility of avoiding expensive data collection efforts from large pools of individuals to train generalized models.

Our results have also shown that personalized predictors substantially outperform their generalized counterparts in approximately 30 percent of the individuals, suggesting the possibility of developing personalized models for sleep-wake states detection that are better able to cope with differences in individual characteristics or sleeping conditions (e.g., ambient noise). Some directions for further research include: development of personalized variants of the recently introduced fully unsupervised models for sleep-wake states prediction and sleep parameter estimation [36], as well as application of the resulting methods to physical activity recognition [27].

REFERENCES

[1] M. Hirshkowitz, "The history of polysomnography: tool of scientific discovery," *Sleep Medicine: A Comprehensive Guide to Its Development, Clinical Milestones, and Advances in Treatment*, pp. 91–100, 2015.

[2] A. Sadeh, "The role and validity of actigraphy in sleep medicine: an update," *Sleep medicine reviews*, vol. 15, no. 4, pp. 259–267, 2011.

[3] D. F. Kripke, D. Mullaney, S. Messin, and V. G. Wyborney, "Wrist actigraphic measures of sleep and rhythms," *Electroencephalography and clinical neurophysiology*, vol. 44, no. 5, pp. 674–676, 1978.

[4] D. F. Kripke, E. K. Hahn, A. P. Grizas, K. H. Wadiak, R. T. Loving, J. S. Poceta, F. F. Shadan, J. W. Cronin, and L. E. Kline, "Wrist actigraphic scoring for sleep laboratory patients: algorithm development," *Journal of sleep research*, vol. 19, no. 4, pp. 612–619, 2010.

[5] D. Mullaney, D. Kripke, and S. Messin, "Wrist-actigraphic estimation of sleep time," *Sleep*, vol. 3, no. 1, pp. 83–92, 1980.

[6] A. Sadeh, J. Alster, D. Urbach, and P. Lavie, "Actigraphically based automatic bedtime sleep-wake scoring: validity and clinical applications," *Journal of Ambulatory Monitoring*, vol. 2, no. 3, pp. 209–216, 1989.

[7] A. Sadeh, M. Sharkey, and M. A. Carskadon, "Activity-based sleep-wake identification: an empirical test of methodological issues," *Sleep*, vol. 17, no. 3, pp. 201–207, 1994.

[8] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, pp. 461–469, 1992.

[9] M. Marino, Y. Li, M. N. Rueschman, J. Winkelman, J. Ellenbogen, J. Solet, H. Dulin, L. F. Berkman, and O. M. Buxton, "Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography," *Sleep*, vol. 36, no. 11, pp. 1747–1755, 2013.

[10] G. Jean-Louis, D. F. Kripke, W. J. Mason, J. A. Elliott, and S. D. Youngstedt, "Sleep estimation from wrist movement quantified by different actigraphic modalities," *Journal of neuroscience methods*, vol. 105, no. 2, pp. 185–191, 2001.

[11] G. Jean-Louis, D. F. Kripke, R. J. Cole, J. D. Assmus, and R. D. Langer, "Sleep detection with an accelerometer actigraph: comparisons with polysomnography," *Physiology & behavior*, vol. 72, no. 1, pp. 21–28, 2001.

[12] L. de Souza, A. A. Benedito-Silva, M. L. N. Pires, D. Poyares, S. Tufik, and H. M. Calil, "Further validation of actigraphy for sleep studies," *Sleep*, vol. 26, no. 1, pp. 81–85, 2003.

[13] J. Hedner, G. Pillar, S. D. Pittman, D. Zou, L. Grote, and D. P. White, "A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients," *Sleep*, vol. 27, no. 8, pp. 1560–1566, 2004.

[14] M. Enomoto, T. Endo, K. Suenaga, N. Miura, Y. Nakano, S. Kohtoh, Y. Taguchi, S. Aritake, S. Higuchi, M. Matsuura, K. Takahashi, K. Mishima, "Newly developed waist actigraphy and its sleep/wake scoring algorithm," *Sleep and Biological Rhythms*, vol. 7, no. 1, pp. 17–22, 2009.

[15] J. Tilmanne, J. Urbain, M. V. Kothare, A. V. Wouwer, and S. V. Kothare, "Algorithms for sleep–wake identification using actigraphy: a comparative study and new results," *Journal of sleep research*, vol. 18, no. 1, pp. 85–98, 2009.

[16] M. L. Lamprecht, A. P. Bradley, T. Tran, A. Boynton, and P. I. Terrill, "Multisite accelerometry for sleep and wake classification in children," *Physiological measurement*, vol. 36, no. 1, p. 133, 2014.

[17] G. Orellana, C. Held, P. Estevez, C. Perez, S. Reyes, C. Algarin, and P. Peirano, "A balanced sleep/wakefulness classification method based on actigraphic data in adolescents," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp. 4188–4191, 2014.

[18] A. Domingues, T. Paiva, and J. M. Sanches, "Sleep and wakefulness state detection in nocturnal actigraphy based on movement information," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 426–434, 2014.

[19] T. Willemen, D. Van Deun, V. Verhaert, M. Vandekerckhove, V. Exadaktylos, J. Verbraecken, S. Van Huffel, B. Haex, and J. Vander Sloten, "An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification," *IEEE journal of biomedical and health informatics*, vol. 18, no. 2, pp. 661–669, 2014.

[20] T. L. Sletten, A. Y. Segal, E. E. Flynn-Evans, S. W. Lockley, and S. M. Rajaratnam, "Inter-individual differences in neurobehavioural impairment following sleep restriction are associated with circadian rhythm phase," *PloS one*, vol. 10, no. 6, p. e0128273, 2015.

[21] T. Rusterholz, L. Tarokh, H. Van Dongen, and P. Achermann, "Interindividual differences in the dynamics of the homeostatic process are trait-like and distinct for sleep versus wakefulness," *Journal of sleep research*, vol. 26, no. 2, pp. 171–178, 2017.

[22] T. T. Dang-Vu, S. M. McKinney, O. M. Buxton, J. M. Solet, and J. M. Ellenbogen, "Spontaneous brain rhythms predict sleep stability in the face of noise," *Current Biology*, vol. 20, no. 15, pp. R626–R627, 2010.

[23] S. M. McKinney, T. T. Dang-Vu, O. M. Buxton, J. M. Solet, and J. M. Ellenbogen, "Covert waking brain activity reveals instantaneous sleep depth," *PLoS One*, vol. 6, no. 3, p. e17351, 2011.

[24] O. M. Buxton, J. M. Ellenbogen, W. Wang, A. Carballeira, S. O'Connor, D. Cooper, A. J. Gordhandas, S. M. McKinney, and J. M. Solet, "Sleep disruption due to hospital noises - a prospective evaluation," *Annals of internal medicine*, vol. 157, no. 3, pp. 170–179, 2012.

[25] O. M. Buxton, M. Pavlova, E. W. Reid, W. Wang, D. C. Simonson, and G. K. Adler, "Sleep restriction for 1 week reduces insulin sensitivity in healthy men," *Diabetes*, vol. 59, no. 9, pp. 2126–2133, 2010.

[26] O. M. Buxton, M. Pavlova, W. Wang, F. L. Scheer, E. B. Klerman, S. P. O'Connor, J. H. Porter, D. T. McLaren, D. T. Cooper, J. M. Ellenbogen, "Examining the effects of daytime eszopiclone administration on daytime sleep and nighttime wakefulness: a randomized, double-blind, placebo-controlled, crossover trial in shift workers," *Sleep*, vol. 36, pp. A184 (Abstract Supplement), 2013.

[27] Y. Zheng, W.-K. Wong, X. Guan, and S. Trost, "Physical activity recognition from accelerometer data using a multi-scale ensemble method." in *IAAI*, 2013.

[28] T. M. Mitchell, "Machine learning," WCB/McGraw-Hill, 1997.

[29] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 215–242, 1958.

[30] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[31] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *European Conf. on computational learning theory*. Springer, pp. 23–37, 1995.

[32] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, pp. 785–794, 2016.

[33] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[34] P. Baldi, S. Brunak, Y. Chauvin, C. AF. Andersen, H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000.

[35] A. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[36] Y. El-Manzalawy, O. M. Buxton, and V. Honavar, "Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering," *IEEE Intl. Conf. on Bioinformatics and Biomedicine*, pp. 718-723, 2017.