Extraction of Major Structural Elements for Successful Aging in Korea through Social Big Data Analysis

Seo-youn Hong

Abstract ***This study aimed to extract key words which can predict 'health-related quality of life' and 'successful aging' among the Korean old people using social big data, and rank those key words in accuracy (Mean Decrease Accuracy (%IncMSE)) and importance (Mean Decrease Gini (IncNodePurity)) in predicting dependent variables. To analyze the data, this study applied Random Forest analysis in R-3.5.0 Version Program. It was found that the relative importance levels (Mean Decrease Gini (IncNodePurity)) of variables were as follows: hobby, preparation, education, sports, volunteer service, love, exercise, welfare, job, and medical care, etc. and, in accuracy levels in predicting successful aging (Mean Decrease Accuracy (%IncMSE)), the rank order of variables were as follows: hobby, love, recognition, sports, welfare, exercise, education, pension, depression, and medical care, etc. In particular, 'hobby' activities of old people showed higher importance and accuracy than those of other word, proving that it is an important variable to predict successful aging among Korean old people.***

Keywords: Successful Aging, Social Big Data, Random Forest Analysis

I. INTRODUCTION

To live a successful old life is important to anybody. People's attention has shifted from simple survival to high-quality life, and the efforts and attention to live successful old life have increased. Reflecting such attention, longevity areas called 'blue zones' across the world have become hot issues. Researchers select 5 blue zones across the world, and, observing lifestyles of old people living there, do researches on the ways of living healthy and happy life in old age [1]. As such, the global attention on the aging society is on healthy and happy life in old age. Accordingly, it is necessary now to find out ways to live happy old life in comprehensive aspect, and find predictive elements of it, and extract its constituent elements.

Academic researches on ways of leading successful aging began in the early 1980s by the American Old Age Sociology Association, and have continued ever since [2]. However, as successful aging has multi-sided characteristics which are different in different ages, cultures, and generations, it is defined in various ways by researchers in different countries. Korea, among the OECD countries, is entering the aging society the most rapidly in Asia. As of 2017, old population in Korea took up 13.8% of the whole population. It is expected that the proportion will reach 41.1% by 2060. Thus, it is urgent to prepare countermeasures on the trend [3].. And, it is not clear that what are universal criteria on predictive elements [4].

Revised Manuscript Received on May 06, 2019

Seo-youn Hong, Soonchunhyang University South, Korea.

Accordingly, the discovery of elements predicting successful aging can be suggested as the possibility to apply to prepare such countermeasures. However, as researchers in different fields offer different sets of predictive elements, there are not sufficient researches on comprehensive constituent elements

Together with 'successful aging', the concept of 'health-related quality of life' is another area of research. Health-related quality of life of old people is not simply their subjective evaluation of health conditions, but a concept with a broader sense. It is a multilateral concept covering physical function, daily life function, social role function, and even subjective happiness [5]. While various elements affect the life of old people, existing researches have focused on specific elements to examine health-related quality of life of them [6]. The problem, however, is that, on the question of what elements affect quality of life, different researchers have different opinions, and there is no agreed-upon formula. So, it is somewhat difficult to approach elements affecting quality of life comprehensively. And, there have been not sufficient researches on how the elements which can be used to predict health-related quality of life are specifically related with successful aging, and what elements should be considered prior to others [7]. Consequently, it is necessary to comprehensively extract elements related with successful aging, and do researches on the relative rank order of the elements predicting successful aging in terms of importance and accuracy.

In the age of the 4th Industrial Revolution, big data is used in various fields to predict future directions and get hints about the future. Big data analysis is done excluding judgement and intention of researchers, and, since enormous amount of data is used to get the outcome, it is possible to overcome the limits of existing researches. Accordingly, it can be meaningful to examine major variables predicting successful aging and relative importance and accuracy of those variables based on big data, and to apply the findings to various researches on the elderly.

Therefore, this study based on big data in Korea which is rapidly changing into a highly aged society, this research intends to discover major variables to predict successful aging, and analyze relative importances and accuracies of those variables, and to use the findings in preparing countermeasures for future related researches.

II. RESEARCH METHODS

2.1. Subjects

This research considered text-based web documents (buzz) which can be collected from blogs, Jisik-in, and news of NAVER web site which

takes up 80% of Internet access by Koreans as social big data, and collected such data. As the key word 'successful aging' is a scholastic term, it has limits to be used as the only key word in NAVER. To collect all the related texts effectively, there was a meeting of specialists in various areas such as big data, welfare for the elderly, leisure for the elderly, and a researcher in the Elderly Happiness Research Center. The meeting selected 14 key words related with successful aging in total.

Finally selected 14 key words were as follows: seonggongjeokin nohu [successful aging], seonggongjeokin nonyeon [successful old years], seonggongjeokin nohusenghwal [successful aging], areumdaun nohusenghwal [beautiful old life], hengbokhan nonyeon [happy old years], hengbokhan nohu [happy old life], japneulneunbeop [way of getting old well], meotitneun nohu [charming old life], 'pumwuiitneun nohu [elegant old life], hengbok & noin [happiness & old person], senghwalmanjok & noin [life satisfaction & old person], hengbokhan sam & noin [happy life & old person], salmeoi hengbok & noin [life satisfaction & old person]. This research searched for texts containing those key words in NAVER from January 2010 to January 2018. Texts were searched for per hour without considering weekend or holiday. The number of collected texts was 2,712 in total: 1,083 blog texts, 475 news articles, and 1,154 articles in jisik-in.

To collect social big data, this research used Crawler, and, from unstructured texts, eliminated repeated texts, and texts containing stopwords. And, selected texts went through the structuring spacing words and rules of spelling and information extraction process.

2.2. Data processing Methods

2.2.1. Focus Group Discussion (FGD)

To explore major predictive variables which affect 'health-related quality of life' of the elderly, this study convened focus group discussions (FGDs) in the forms of specialist panels. FGDs were held from September to November, 2017. They consisted of 10 groups, with each group consisting of 4-10 members. In a typical FGD, the moderator would raise a theme, and participants discussed the issue freely for about 1 hour to 1 and half hours. Unlike individual in-depth interview, FGD consisting of various people allowed participants to express various kinds of thoughts, experiences, and opinions. And, while exchanging opinions among participants, they produced key words essential to 'healthy quality of life'. The important key words were different in different groups. Based on the finally extracted key words and degrees of importance, the meeting of specialists in various areas like sports medical science, senior welfare, public health, and big data determined 32 predictive variables on 'health-related quality of life'.

2.2.2. Methods

During the period from January 2010 to January 2018, this study crawled all the Web pages containing 32 words which had been determined as predictive variables and 14 key words related with 'successful aging' from blogs, Jisikin, and news web sites in the portal site 'Naver'. In total, 7,002 web pages including 1,178 blogs, 3,508 Jisikins, and 2,316 news were searched. In order to examine how the target variable 'successful aging' can be explained by 'health-related quality of life' predictive variables, the data was under Random Forest analysis using the R-3.5.0 Version Program. Each of the 7,002 cases of web pages was under the coding process. If any of the 32 predictive variables on 'health-related quality of life' or 14 key words related with 'successful aging' is included in the contents of the page, value 1 was given, and if it is not, value 0 was given. Then, Random Forest analysis, a machine learning technique, was done to extract accuracy (Mean Decrease Accuracy (%IncMSE)) and importance (Mean Decrease Gini (IncNodePurity)) of variables predicting successful aging.

III. RESULTS

3.1. Extraction of 'Health-Related Quality of Life' Predictive Variables

To extract major variables which can predict health-related life quality, this research performed focus group discussion (FGD) consisting of specialists in various fields such as public health, medical science, insurance for the elderly, welfare for the elderly, exercise for the elderly, leisure for the elderly, and sports medicine for the elderly. Participants the FGD in the first round offered and exchanged their ideas, and 102 key words on health-related quality of life emerged from the discussion. Then, key words sharing the same meaning were combined, and, in the second round, there was an interpretation process of deciding which key word is related with which key word, and the key words were ranked depending on relative importance of them. Based on the selected key words and rank order among them, the meeting of specialists - 1 gerontologist, 1 old welfare specialist, 1 public health specialist, and 1 big data specialist - determined 30 variables predicting 'health-related quality of life. Through this process, variables were divided into two layers: big domains and middle domains.

There are 12 big domains: relationship, society, education, religion, economy, disease, leisure, environment, body, activity, mentality, and spirit. Each of these big domains is divided into some middle domains. For example, relationship is divided into family, friend, couple, and offspring; society into welfare; economy into medical care, preparation, insurance, pension, property and house; leisure into travel and sports; body into exercise and sleep; activity into volunteer service, challenge, hobby and job; mentality into love, recognition, achievement, depression, emotion, panic and bipolar disorder; spirit into spiritual, and mind. Organization of predictive variables is shown in Table 1.

Table 1: Extraction of 'health-Related Quality of Life' predictive variables

No	Big domain	Middle domain	N
1	Relationship	Family	2473
		Friends	3416
		Couple	1404
		Offspring	1253
2	Society	Welfare	4269
3	Education	Education	8404
4	Religion	Religion	7967
5	Economy	Medical care	3825
		Preparation	2897
		Insurance	11818
		Pension	3002
		Property	3316
		House	1298
6	Disease	Disease	1681
7	Leisure	Travel	1686
		Sports	1649
8	Environment	Environment	4934
9	Body	Exercise	4242

		Sleep	301
10	Activity	Volunteer Service	3039
		Challenge	1139
		Hobby	374
		Job	1584
11	Mentality	Love	9184
		Recognition	3163
		Achievement	813
		Depression	369
		Emotion	283
		Panic	91
		Bipolar Disorder	45
12	Spirit	Spiritual	2154
		Mind	4712

3.2. Importance (IncNodePurity) of Variables Predicting Successful Aging

In Random Forest analysis, Mean Decrease Gini (IncNodePurity) relates to the loss function by which best splits are chosen [8]. It was found that, among 32 predictive variables affecting 'health-related quality of life', Mean Decrease Gini of 'hobby' was the highest. It means that, if 'hobby' is excluded, predictive accuracy of Random Forest greatly falls down. The words with next highest Mean Decrease Ginis were preparation, education, sports, volunteer service, love, exercise, welfare, job, medical care, family, environment, pension, depression, mind, couple, insurance, disease, travel, friends, recognition, house, offspring, challenge, property, religion, achievement, emotion, sleep, spiritual, panic and bipolar disorder in descending order.
Specifically, Mean Decrease Gini (IncNodePurity) of hobby was 36.347, the highest among variables. Mean Decrease Ginis of key words, preparation, education, sports, volunteer service and love were about 24, lower than that of hobby. All the predictive variables and rank orders of them are shown in Table 2.

Table 2: Importance (IncNodePurity) of variables predicting successful aging ranking

No	Keyword	IncNodePurity
1	Hobby	36.347
2	Preparation	24.903
3	Education	24.893
4	Sports	24.775
5	Volunteer Service	24.714
6	Love	24.676
7	Exercise	23.898
8	Welfare	23.287
9	Job	23.055
10	Medical Care	22.838
11	Family	22.452
12	Environment	22.279
13	Pension	21.499
14	Depression	21.245
15	Mind	21.162
16	Couple	20.817
17	Insurance	20.424
18	Disease	20.126
19	Travel	19.838
20	Friends	19.802
21	Recognition	18.899
22	House	18.805
23	Offspring	18.800
24	Challenge	17.940
25	Property	17.473
26	Religion	17.047
27	Achievement	12.899
28	Emotion	10.841
29	Sleep	9.780
30	Spiritual	3.455
31	Panic	3.106
32	Bipolar Disorder	1.305

3.3. Accuracy (%IncMSE) of Variables Predicting Successful Aging

In random forest analysis, Mean Decrease Accuracy (%IncMSE) is the most robust and informative measure [8]. Among the variables predicting health-related quality of life which influence successful aging, Mean Decrease Accuracy of variable ‘hobby’ was the highest. That is, Mean Decrease Accuracy of hobby in predicting successful aging was shown to be 49.910%, which means that, only with the key word ‘hobby’, successful aging can be predicted with the 49.910% probability. The rank order of accuracy of variables predicting successful aging was as follows in descending order: love, recognition, sports, welfare, exercise, education, pension, depression, medical care, job, volunteer service, couple, preparation, offspring, house, disease, insurance, mind, travel, property, family, panic, challenge, religion, environment, sleep, achievement, bipolar disorder, friend, emotion, and spiritual. As the case of Mean Decrease Gini, ‘hobby’ showed the highest accuracy (49.910%) in predicting successful aging, followed by that of ‘love’ (40.555%). The accuracies of three variables, ‘hobby’, ‘love’ and 'recognition' were conspicuously higher than those of other variables. In contrast, the accuracy of variable ‘spiritual’ was 1.783%, the lowest among those variables. All the variables and the rank order of them are shown in Table 3.

Table 3: Accuracy (%IncMSE) of variables predicting successful aging ranking

No	Keyword	%IncMSE
1	Hobby	49.910
2	Love	40.555
3	Recognition	40.207
4	Sports	37.624
5	Welfare	36.015
6	Exercise	34.475
7	Education	34.365
8	Pension	34.252
9	Depression	33.481
10	Medical Care	31.150
11	Job	31.130
12	Volunteer Service	30.437
13	Couple	30.012
14	Preparation	29.075
15	Offspring	25.234
16	House	24.671
17	Disease	24.649
18	Insurance	23.625
19	Mind	22.546
20	Travel	22.116
21	Property	20.344
22	Family	18.671
23	Panic	17.749
24	Challenge	17.517
25	Religion	17.338
26	Environment	16.700
27	Sleep	16.112
28	Achievement	14.529
29	Bipolar Disorder	13.177
30	Friend	12.543
31	Emotion	10.835
32	Spiritual	1.783

And, Figure 1 shows the diagram of Random Forest analysis of predictive variables.

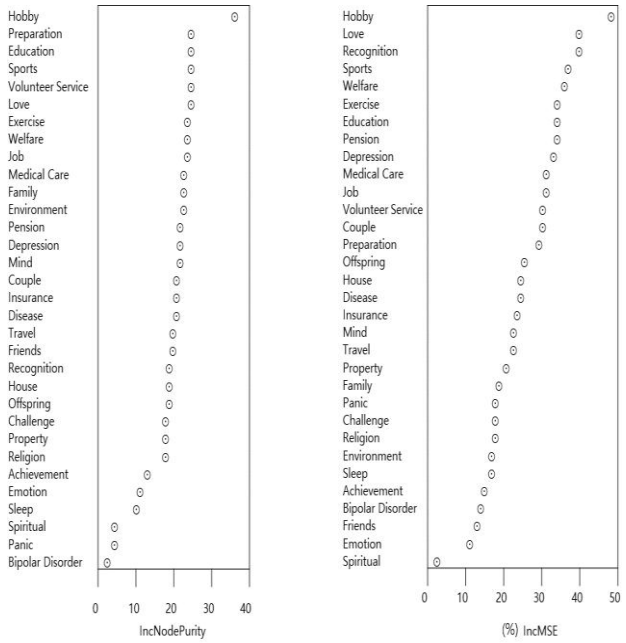

Fig. 1: Random Forest analysis of predictive variables

IV. Conclusion

This research extracted important key words which can predict successful aging from social big data in Korea. Based on extracted key words, this research performed Random Forest machine learning, one of data mining techniques. The aim of this research was to find rank orders in Mean Decrease Gini (IncNodePurity) and Mean Decrease Accuracy (%IncMSE) of variables in predicting successful aging.

What is remarkable in the findings of the analysis is that variable 'hobby' was the highest in Mean Decrease Gini (IncNodePurity) and Mean Decrease Accuracy (%IncMSE) among all the key variables.

It means that whether the elderly enjoy their hobbies, and how much they participate in hobby activities are important predictors on whether they lead successful aging. Accordingly, future researches on successful aging need to pay attention to the variable 'hobby', and other variables related with hobby. It is necessary to do further researches on how hobby can be applied in welfare programs for the elderly and education for them.

And, it was found that 'love' was ranked 6 in Mean Decrease Gini, and 'love' and 'recognition' showed over 40% in Mean Decrease Accuracy. Such findings can also be identified in a meta-analysis integrating all the variables related with successful aging. The analysis showed strong effects of such psychological variables [4]. That is, we can identify that, to Korean old people, positive psychological variables are important ones to predict successful aging. Accordingly, it is necessary to find out how positive mentality like 'love' and 'recognition' can be applied to old education and old welfare programs.

Besides, key words related with negative mentality like depression, panic and bipolar disorder were also extracted. Especially, depression of the elderly in Korea has continued to attract attention [9]. In this research as well, depression showed rather high importance level and accuracy level in predicting successful aging. Thus, there should be means to intervene depression of the elderly. This research also found out that sports and exercise showed rather high importance levels and accuracy levels. Exercise has proved to be an effective intervention means to reduce depression among old people [10]. Considering mutual relations among key words in affecting successful aging, it is necessary to do additional researches on ways to intervene depression among old people.

Rowe and Kahn (1987), early researchers on successful aging, offered prevention of disease and disability, maintenance of physical and cognitive functions, and active involvement in social activities as core concepts of successful aging. However, as life conditions and living standards have improved in many countries, physical health of old people have improved, and they can enjoy material abundance. Some scholars have suggested that, rather than examining and prohibiting risk factors for improvement of life quality of the elderly, it is necessary to develop protective factors to protect the old. Afterwards, many scholars have suggested importance of 'positive spirituality' as potential resource for successful aging. World Health Organization included spiritual goodness among important elements of health [11] [12] [13].

Those specialists this study invited to discuss in specialist meeting understood the importance of spirituality, and chose it as a key word. But, Random Forest analysis showed low importance and accuracy of it. However, this finding is rather low influence among 30 key words affecting successful aging. It is necessary to do additional researches on it. That is, this research has limits in the sense that it approached the issue only using key words. So, based on the findings of this research, it is necessary to do demand survey to the elderly and to specialists. And it seems that the variable 'spiritual' should not be limited to have religious connotation, but should be interpreted as a comprehensive concept which provides people with motive to pursue meaning and goal in individual life.

For successful aging, social activities and social relations of the elderly are important [4][14]. In this study as well, we can identify key words related with social relations such as volunteer service, job, and friends, Such social relations of the elderly have been continuously treated [15][16]. Such social relations are highly related with psychological characteristic of them such as depression, confidence, achievement, challenge, and recognition [17]. Future researches need to use the findings of this research and examine the interrelationship among variables, and search for ways to improve social relations of the elderly through leisure activities, sports, hobby-related activities, and volunteer service activities.

Researches related with successful aging in Korea have commonly been done focusing on the variables researchers are interested in. Thus, they have limits in the sense that there was no comprehensive research on all the variables, and that there has not been additional research based on the comprehensive study. This study has limits that it is based on social big data only in Korea, and extracted predictive variables as key words. However, it is meaningful in the sense that, it generated findings in a rapidly

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

aging society in Asia among OECD countries, and comprehensively extracted key words from a huge amount of social big data. In the future study, it is necessary to examine how those key words are related with each other. It is also necessary to do demand surveys to the elderly and specialists on the elderly, based on key words extracted in this study. By doing that, we can seek concrete ideas on education for the old, welfare programs for the old, and marketing targeting the old, etc.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea.
(NRF-2017S1A5B5A02025639)

REFERENCES

1. D. Buettner, *The blue zones: 9 lessons for living longer from the people who've lived the longest*, National Geographic Books, (2012).
2. J. W. Rowe & R. L. Kahn, "Human Aging: Usual and Successful", *Science*, Vol.237, pp. 143-149, (1987).
3. Statistics Korea, 2017 Aged Statistics, (2017).
4. Y. J. Jeong, N. Y. Yu, B. A. Kim, H. J. Shin & Y. S. Chong, "A Meta-Analysis on Variables Related with Successful Aging", *Journal of the Korean Gerontological Society*, Vol.34, No.4, pp.781-797, (2014).
5. G. I. Kempen, J. M. Ormel, E. I. Brilman, & J. Relyveld, "Adaptive response among Dutch elderly: the impact of eight chronic medical conditions on health -related quality of life", *The Australian journal of politics and history*, Vol.87, No.1, pp.38-44, (1997).
6. S. M. Moon, "Gender differences in the impact of socioeconomic, health-related, and health behavioral factors on the health-related quality of life of the Korean elderly", Journal of digital Convergence, Vol.15, No.6, pp. 259-271, (2017).
7. S. Y. Hong, "Perspective on Aging Using Random Forest Analysis: Social Big Data Approach", International Journal of Eldery Welfare Promotion and Management, Vol.2, No.2, pp. 61-66, (2018).
8. T. M. Song & J. Y. Song, *Social Big Data and Future Prediction with Machine Learning*, Hannarae, (2017).
9. D. B. Kim & E. S. Sohn, "A Meta-analysis of the Variables Related to Depression in Elderly", *Journal of the Korean Gerontological Society*, Vol.25, No.4, pp. 167-187, (2005).
10. J. J. Bae & H. C. Park, "The Effect of Exercise on Senior Depression in Local Area", *Korean Journal of Physical Education*, Vol.53, No.3, pp. 549-557, (2014).
11. S. H. Chung & J. H. Lim, "Exploring the links between spirituality and successful aging of older adults", *The Korean Journal of the Human Development*, Vol.20, No.4, pp. 59-79, (2013).
12. M. R. Crowther, M. W. Parker, A. W. Achenbaum, W. L. Larimore & H. G. Koenig, "Rowe and Kahn's model of successful aging revisited: Positive spirituality-the forgatten factor", *The Gerontologist*, Vol.42, No.5, pp. 613-620, (2002).
13. E. Sandler & S. Biffs, "Exploring the links between spirituality and successful aging", *Journal of Social Work Practice*, Vol.20, No.3, pp. 267-280, (2006).
14. R. Crosnoe & G. H. Elder, "Successful adaptation in the later life", *International Journal of Aging and Human Development*, Vol.65, No.4, pp. 309-328, (2002).
15. R. C. Mannell & S. Dupuis, *Life Satisfaction, Encyclopedia of Gerontology*, New York, (1996).
16. N. Morrow-Howell, *Productive Engagement of Older Adults: Effects on Well-Being*, Center for Social Development, Washington University, (2000).
17. S. I. Kim, "Relationship among Social Support, Psychological Well-Being and Depression of Elderly Participate in Physical Activity", *Korean Journal of Physical Education*, Vol.51, No.1, pp. 333-344, (2012).