# Decomposition Analysis of Index of Multiple Deprivation (IMD) Based on Shapley Value

Candidate:Xiaohan Feng

Submission Date: xxxx/xx/xx

Module Name: MSc Dissertation

Module ID: CASA0004

Supervisor: Prof. Huanfa Chen

Word Count: xxxxxx

This dissertation is submitted in part requirement for the MSc in the Centre

for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

Contents

[Research question 3](#_Toc74834369)

[Literature Review 3](#_Toc74834370)

[IMD 3](#_Toc74834371)

[The definition of IMD 3](#_Toc74834372)

[The function of IMD 3](#_Toc74834373)

[The limitation of IMD 3](#_Toc74834374)

[Limitation about multiple indexes 3](#_Toc74834375)

[Limitation about subjectivity 4](#_Toc74834376)

[Limitation about data 4](#_Toc74834377)

[Shapley Value 4](#_Toc74834378)

[Shapley Value in geography / GIS / urban research 5](#_Toc74834379)

[The regional policy suggestion of the CO2 emissions by Shapley value 5](#_Toc74834380)

[Other fields using Shapley Value related to urban research 5](#_Toc74834381)

[Methodology 6](#_Toc74834382)

[Shapley Value 6](#_Toc74834383)

[Background 6](#_Toc74834384)

[properties 6](#_Toc74834385)

[SHAP(SHapley Additive exPlanations) and KernelSHAP 7](#_Toc74834386)

[Reference 8](#_Toc74834387)

## Research question

how do the indices from 7 domains of deprivation contribute to the individual and overall values of Index of Multiple Deprivation (IMD) in 2019 for small areas (Lower-layer Super Output Area) across England by the Shapley value based on public data from National Statistics?

Based on the result we get (knowing which indicators contribute most), what corresponding measurement could be taken for policymakers to alleviate the problem brought by the high IMD score and improve the overall living circumstances in areas that have a high rank of IMD?

## Literature Review

### IMD

#### The definition of IMD

It is an interesting area to quantify the spatial variation of social and economic circumstances of different areas. One of the most famous attempts is the English government’s Index of Multiple Deprivation (IMD), which measures deprivation in England locally. It is made up of seven domain measures of deprivation, including Income Deprivation; Employment Deprivation; Education, Skills and Training Deprivation; Health Deprivation and Disability; Crime; Barriers to Housing and Services; Living Environment Deprivation (National Statistics, 2019). The main domains of IMD are similar across the world. For example, the IMD of New Zealand is similar to the IMD of England(Exeter *et al.*, 2017). The IMD of German has five domains similar to the IMD of England (Maier, 2017). In fact, the theoretical basis of determining deprivation is taken from Townsend’s work in the 1980s, in which he defined the deprivation as lacking “diet, clothing, housing, household facilities, and fuel and environmental, educational, working and social conditions, activities and facilities” (Noble *et al.*, (2006) quoting Townsend, p. 172).

#### The function of IMD

IMD can be used in many fields. Government can use IMD to help target resources to “priority areas”. In other words, it is a useful tool to guide resource distribution. It was ever used to help the central government of England to determine the eligible amount of Neighbourhood Renewal Fund monies to the local authorities (Cabinet Office, 2001). And some government framework for regeneration and funds allocation documents made explicit references to IMD (Communities and Local Government, 2008; Communities and Local Government, 2009). IMD is also broadly used as a key indicator to identify the need or deprivation of local people, within academic circles (Cento Bull and Jones, 2006; Kintrea, 2007; Macintyre, Macdonald and Ellaway, 2008).

#### The limitation of IMD

Given the fact that IMD is more and more widely employed in many fields, it is necessary to understand the limitation of it.

##### Limitation about multiple indexes

IMD is the combination of multiple domains, thus, it shows an overview of the local deprivation. However, on the other hand, it may also be difficult to eliminate the measurement error and aggregation irrationality of so many domains of data. And as an aggregated index, it suffers that “deficits in some sectors, which actually threaten the health of the whole system” (Bossel, 1999). The relationship of the sub-index inside IMD may also cause the problem of estimate the deprivation. For example, poor health may affect employment opportunities and income, which further affect the Barriers to Housing and Services (Briggs, Abellan and Fecht, 2008).

##### Limitation about subjectivity

What’s more, due to lots of subjective intervene, IMD always raises concerns about paternalism. For example, the allocation of weights to different domains, the choice of indicators in each domain, and many other mechanisms are all determined by the expert’s judgement (Watson et al., 2019); the exponential transformation on the domains is done without explanation and implication (Deas et al., 2003); the process to construct the multiple deprivations lacks the general theory support (Clelland and Hill, 2019). Therefore, the validity of the results and the utility in the policy decisions may face the challenge.

##### Limitation about data

Some of the individual indicators used in the computation of the local IMD were collected at a district or national scale. Without using the exact ward data, the estimated value may distort the accuracy of the local IMD (Deas *et al.*, 2003). Some indicators are double-counted across more than one domain. Such as Severe Disablement Allowance contributes to not only the Employment Deprivation Domain but also the Health Deprivation and Disability Domain. Though some expert argued that it is legitimate to count the indicators as two deprivations (University of Oxford, 2000), it still raises concerns about the veracity especially when we do the overall factor analysis of all the indicators of IMD (Deas et al., 2003).

### Shapley Value

Shapley Value aims to figure out the reasonably expected payoff for each player in a cooperative game. One of the most famous problems solved by Shapley Value is the cost-sharing problem(Gul, 1989; Pérez-Castrillo and Wettstein, 2001; CHUN, HU and YEH, 2017). For example, Siano, Gallo and Glielmo (2015) calculate the distribution of costs of the shared travel for each person based on the Shapley value after he presented a demand and supply managing algorithm for the shared transportation system, which is similar to the urban railway ticket pricing mechanism designed by Lu *et al.* (2010) using Shapley. Along with the development of Machine Learning models, Shapley Value is popular to be applied to helping understand the intrinsic influence factors of Machine Learning models(Lundberg and Lee, 2017; Lundberg, Erion and Lee, 2018), since Machine Learning models are always the “black box” which conceals the importance of the factors for us. Smith and Alvarez (2021) used Shapley Value to observe the marginal impact of each mortality factor in his Machine Learning model for a case of COVID19, which is helpful to detect anomalous patterns when treating patients. Petrosjan and Zaccour (2003) studied the time-consistent Shapley Value to allocate the cost of pollution reduction for countries. It was proved that using Shapley outcomes, each country can receive a fair time-consistent cost and the total cost would be lower than the sum of the cost of each country assuming playing a noncooperative game. Reddy, Shevkoplyas and Zaccour (2013) also discussed Shapley Value over time. This theory for dynamic stochastic games can be widely applied in many fields to help build a long-term cooperation relationship among players.

#### Shapley Value in geography / GIS / urban research

##### The regional policy suggestion of the CO2 emissions by Shapley value

Many experts in China using the Shapley Value decomposition method to analyze the key factors of CO2 emissions in different areas(Yu, Wei and Wang, 2014; Liang *et al.*, 2018). The method helped them find similar factors, which contribute most to the CO2 emissions in China’s thermal electricity generation and European agriculture (Li *et al.*, 2016; Yan *et al.*, 2018). Zhang, Wang and Dac (2014) compared the entropy and Shapley Value methods in allocation the carbon quota to different regions in China. He found the result of the entropy method has a positive effect on the task of the Shapley value method. The final analysis results of Shapley Value let him come up with some policy implications for the CO2 emissions reduction in China regionally. Wen and Hao (2020) combined the Shapley Value with Spectral Clustering algorithm to decompose the factors of CO2 emissions at the provincial level in China. He found carbon intensity played a significant role in most provinces, but there are other factors that affect the CO2 emissions differently in different provinces. Based on the PSO-FCM clustering method and Shapley Value, Yu, Wei and Wang (2014) clustered the 30 provinces of China into four classes according to 13 macro factors which may influence CO2 emissions. He proposed three-parts CO2 emissions reduction strategies to be suitably used at the provincial level. By Shapley Value, he also found the main approaches to reduce the CO2 emissions for each class, respectively. The Shapley Value decomposition method can give customized policy suggestions that vary from region to region in the light of local conditions.

##### Other fields using Shapley Value related to urban research

The Shapley value method is also used in many other fields to provide the intrinsic view about the influence factors on the problem. Chen and Li (2014) used Shapley Value to decompose the income inequality into education disparity, household registration, geographic location, type of job, and gender. Geographic location plays the third important role in the analysis, and the regional difference may also affect other influence factors of income inequality. Aristondo and Onaindia (2020) applied the Shapley Value decomposition method to explore overall poverty change in terms of three poverty components’ changes. He found that the incidence, intensity, and inequality should all be taken into account as the poverty measures according to Shapley Value. He applied the theory to 28 European countries and got the regional results for each country. Dong *et al.*, (2020) employed the regression-based Shapley decomposition method to aid the analysis of haze pollution by quantile regression. The population density of regional inequality is found by Shapley to be the most important factor of regional differences in haze pollution.

## Methodology

### Shapley Value

Shapley value is created by Shapley (1953) from cooperative game theory and aims to assign player’s payout according to their contribution. It is similar to other explanation methods in some cases. However, these methods are based on the certain assumption (take LIME for example, it assumes that the target black-box model can be locally approximated by some interpretable model like decision trees and linear models) and do not have the theory to support it; while Shapley value has the solid theory as it is the only method that satisfies the properties of a fair payout: Efficiency, Symmetry, Dummy, and Additivity (Molnar, 2021). Following is a detailed explanation of these properties.

##### Background

We assume there are total n people in collaboration and they create a profit of What need to do is fairly distribute this profit.

and is the vector of feature values of the instance (player ).

is a subset of the . represents the value generated by the cooperation of the elements in the S. The final Shapley value for player i is .

##### properties

1. **Efficiency:** the full yield of the game is distributed to the players, which means the sum of the Shapley value among players equal to the total value.
2. **Symmetry**: i and j are interchangeable relative to if the contributions of them are equal to all possible coalitions. For all that contains neither i nor j, if , then . That is to say, interchangeable (equivalent) players receive the same amount of payments.
3. **Dummy**: player i is dummy if his contribution to any coalition is zero. That would mean, . For all S, if i is dummy, he should have 0 Shapley value(receive nothing)，.
4. **Additivity**：if a game can be separated into two parts , then the distributed gains from value function should correspond to the gains derived from and the gains derived from . For every coalition S and player i, where the game is defined by This is a prime requisite if a researcher intends to design an evaluation scheme that would be applied to the “systems of interdependent games”(Kuhn, 1997).

The Shapley value is proved by Shapley (1953) to be the only map from the set of all games to payoff vectors that satisfies all four properties.

From the formula above, we can get that the Shapley value of a feature value is its average marginal contribution over all possible coalitions to the payout

### SHAP(SHapley Additive exPlanations) and KernelSHAP

因为我们直接计算shapley的值，不通过模型，所以方法论里就不详细描述其他的比如TreeSHAP这些嘛？

## Reference

*A New Commitment to Neighbourhood Renewal*. (2001). London. Available at: http://www.bris.ac.uk/poverty/downloads/keyofficialdocuments/Neighbourhood Renewal National Strategy Report.pdf.

Aristondo, O. and Onaindia, E. (2020). ‘On measuring the sources of changes in poverty using the Shapley method. An application to Europe’. *Fuzzy Sets and Systems*. Elsevier B.V., 383, pp. 80–91. doi: 10.1016/j.fss.2018.12.011.

Bossel, H. (1999). *Indicators for sustainable development: theory, method, applications*. International Institute for Sustainable Development Winnipeg.

Briggs, D., Abellan, J. J. and Fecht, D. (2008). ‘Environmental inequity in England: Small area associations between socio-economic status and environmental pollution’. *Social Science and Medicine*. Pergamon, 67 (10), pp. 1612–1629. doi: 10.1016/j.socscimed.2008.06.040.

Cento Bull, A. and Jones, B. (2006). ‘Governance and Social Capital in Urban Regeneration: A Comparison between Bristol and Naples’. *Urban Studies*, 43 (4), pp. 767–786. doi: 10.1080/00420980600597558.

Chen, C. and Li, S. (2014). ‘Market transition and income inequality in urban China: Evidence from shapley value decomposition’. *Frontiers of Economics in China*, 9 (2), pp. 309–337. doi: 10.3868/s060-003-014-0016-9.

CHUN, Y., HU, C.-C. and YEH, C.-H. (2017). ‘A Strategic Implementation of the Shapley Value for the Nested Cost-Sharing Problem’. *Journal of Public Economic Theory*, 19 (1), pp. 219–233. doi: 10.1111/jpet.12190.

Clelland, D. and Hill, C. (2019). ‘Deprivation, policy and rurality: The limitations and applications of area-based deprivation indices in Scotland’. *Local Economy: The Journal of the Local Economy Policy Unit*, 34 (1), pp. 33–50. doi: 10.1177/0269094219827893.

Deas, I., Robson, B., Wong, C. and Bradford, M. (2003). ‘Measuring neighbourhood deprivation: A critique of the Index of Multiple Deprivation’. *Environment and Planning C: Government and Policy*, 21 (6), pp. 883–903. doi: 10.1068/c0240.

Dong, F., Yu, B., Pan, Y. and Hua, Y. (2020). ‘What contributes to the regional inequality of haze pollution in China? Evidence from quantile regression and Shapley value decomposition’. *Environmental Science and Pollution Research*. Environmental Science and Pollution Research, 27 (14), pp. 17093–17108. doi: 10.1007/s11356-020-07929-8.

Exeter, D. J., Zhao, J., Crengle, S., Lee, A. and Browne, M. (2017). ‘The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand’. *PLoS ONE*, 12 (8), pp. 1–19. doi: 10.1371/journal.pone.0181260.

Gul, F. (1989). ‘Bargaining foundations of Shapley value’. *Econometrica: Journal of the Econometric Society*. JSTOR, pp. 81–95.

Kintrea, K. (2007). ‘Policies and Programmes for Disadvantaged Neighbourhoods: Recent English Experience’. *Housing Studies*, 22 (2), pp. 261–282. doi: 10.1080/02673030601132920.

Kuhn, H. W. (1997). *Classics in Game Theory*. Princeton University Press (Frontiers of Economic Research). Available at: https://books.google.co.uk/books?id=HyTpw6H5syUC.

Li, T., Baležentis, T., Makutėnienė, D., Streimikiene, D. and Kriščiukaitienė, I. (2016). ‘Energy-related CO2 emission in European Union agriculture: Driving forces and possibilities for reduction’. *Applied Energy*, 180, pp. 682–694. doi: 10.1016/j.apenergy.2016.08.031.

Liang, Y., Niu, D., Zhou, W. and Fan, Y. (2018). ‘Decomposition Analysis of Carbon Emissions from Energy Consumption in Beijing-Tianjin-Hebei, China: A Weighted-Combination Model Based on Logarithmic Mean Divisia Index and Shapley Value’. *Sustainability*, 10 (7), p. 2535. doi: 10.3390/su10072535.

Lu, W., Chen, Y., Zhang, N. and Shi, Y. (2010). ‘Ticket pricing of urban rail transit based on Shapley value method’. *Proceedings - 2010 2nd IEEE International Conference on Information and Financial Engineering, ICIFE 2010*. IEEE, pp. 127–130. doi: 10.1109/ICIFE.2010.5609267.

Lundberg, S. and Lee, S.-I. (2017). ‘A unified approach to interpreting model predictions’. *arXiv preprint arXiv:1705.07874*.

Lundberg, S. M., Erion, G. G. and Lee, S.-I. (2018). ‘Consistent individualized feature attribution for tree ensembles’. *arXiv preprint arXiv:1802.03888*.

Macintyre, S., Macdonald, L. and Ellaway, A. (2008). ‘Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland’. *Social Science & Medicine*, 67 (6), pp. 900–914. doi: 10.1016/j.socscimed.2008.05.029.

Maier, W. (2017). ‘Indizes Multipler Deprivation zur Analyse regionaler Gesundheitsunterschiede in Deutschland’. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 60 (12), pp. 1403–1412. doi: 10.1007/s00103-017-2646-2.

Molnar, C. (2021). *5.9 Shapley Values | Interpretable Machine Learning*. Available at: https://christophm.github.io/interpretable-ml-book/shapley.html#general-idea (Accessed: 15 June 2021).

National Statistics. (2019). *The English Indices of Deprivation 2019*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/833947/IoD2019\_Research\_Report.pdf.

Noble, M., Wright, G., Smith, G. and Dibben, C. (2006). ‘Measuring Multiple Deprivation at the Small-Area Level’. *Environment and Planning A: Economy and Space*, 38 (1), pp. 169–185. doi: 10.1068/a37168.

Pérez-Castrillo, D. and Wettstein, D. (2001). ‘Bidding for the surplus: a non-cooperative approach to the Shapley value’. *Journal of economic theory*. Elsevier, 100 (2), pp. 274–294.

Petrosjan, L. and Zaccour, G. (2003). ‘Time-consistent Shapley value allocation of pollution cost reduction’. *Journal of Economic Dynamics and Control*, 27 (3), pp. 381–398. doi: 10.1016/S0165-1889(01)00053-7.

Reddy, P. V., Shevkoplyas, E. and Zaccour, G. (2013). ‘Time-consistent Shapley value for games played over event trees’. *Automatica*, 49 (6), pp. 1521–1527. doi: 10.1016/j.automatica.2013.02.029.

Shapley, L. S. (1953). ‘"A Value for n% person Games. Contributions to the Theory of Games’. *Annals of Mathematics Studies*, 28 (2), pp. 307–317.

Siano, G., Gallo, M. and Glielmo, L. (2015). ‘A method for managing transportation requests and subdivision costs in shared mobility systems’. *VEHITS 2015 - Proceedings of the 1st International Conference on Vehicle Technology and Intelligent Transport Systems*, pp. 152–158. doi: 10.5220/0005459501520158.

Smith, M. and Alvarez, F. (2021). ‘Identifying mortality factors from Machine Learning using Shapley values – a case of COVID19’. *Expert Systems with Applications*, 176, p. 114832. doi: 10.1016/j.eswa.2021.114832.

*Transforming places; changing lives: A framework for regeneration*. (2008). London.

*Transforming Places; Changing Lives: Taking Forward the Regeneration Framework Department of Communities and Local Government*. (2009). London.

University of Oxford. (2000). *Response to the Formal Consultations on the Indices of Deprivation 2000*.

Watson, V., Dibben, C., Cox, M., Atherton, I., Sutton, M. and Ryan, M. (2019). ‘Testing the Expert Based Weights Used in the UK’s Index of Multiple Deprivation (IMD) Against Three Preference-Based Methods’. *Social Indicators Research*. Springer Netherlands, 144 (3), pp. 1055–1074. doi: 10.1007/s11205-018-02054-z.

Wen, L. and Hao, Y. (2020). ‘Factor decomposition and clustering analysis of CO2 emissions from China’s power industry based on Shapley value’. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*. Taylor & Francis, 00 (00), pp. 1–17. doi: 10.1080/15567036.2020.1776795.

Yan, Q., Wang, Y., Baležentis, T., Sun, Y. and Streimikiene, D. (2018). ‘Energy-Related CO2 Emission in China’s Provincial Thermal Electricity Generation: Driving Factors and Possibilities for Abatement’. *Energies*, 11 (5), p. 1096. doi: 10.3390/en11051096.

Yu, S., Wei, Y. M. and Wang, K. (2014). ‘Provincial allocation of carbon emission reduction targets in China: An approach based on improved fuzzy cluster and Shapley value decomposition’. *Energy Policy*. Elsevier, 66 (2014), pp. 630–644. doi: 10.1016/j.enpol.2013.11.025.

Zhang, Y. J., Wang, A. D. and Da, Y. Bin. (2014). ‘Regional allocation of carbon emission quotas in China: Evidence from the Shapley value method’. *Energy Policy*. Elsevier, 74 (C), pp. 454–464. doi: 10.1016/j.enpol.2014.08.006.