

世界各国のがん死亡率の格差の解析
－サポートベクターマシンによる決定要因の探索－

鈴木孝弘・田辺和俊

Analysis of Cancer Mortality Rates of Many Countries
－ Search of Determinants for Cancers
Using Support Vector Machine －

Takahiro SUZUKI, Kazutoshi TANABE

東洋大学紀要 自然科学篇 第 59 号 抜刷

Reprinted from

Journal of Toyo University, Natural Science

No. 59, pp.73 ~ 87, March, 2015

Tokyo, Japan

世界各国のがん死亡率の格差の解析 ーサポートベクターマシンによる決定要因の探索ー

鈴木孝弘^a・田辺和俊^b

Analysis of Cancer Mortality Rates of Many Countries – Search of Determinants for Cancers Using Support Vector Machine –

Takahiro SUZUKI^a, Kazutoshi TANABE^b

Abstract

Social determinants of cancer mortality attract much attention worldwide, and several studies on statistical analysis of determinants using linear regression method (OLS) have been reported. In this study, a large-scale experiment has been done to analyze determinants of cancer mortality rates of 147 countries using 40 explanatory variables of 5 fields and a non-linear regression support vector machine (SVM) technique. Except for five developing countries, the resulting SVM model shows that just 14 factors represent cancer mortality rates of 142 countries with a higher accuracy of determination coefficient of 0.730 than previous works. It also demonstrates that the degree of national development such as intelligence quotient and health expenditures most significantly contributes to cancer mortality rates.

Keywords: cancer mortality; determinant analysis; support vector machine

1. はじめに

がんは、一般的に悪性腫瘍あるいは悪性新生物と同義であり、皮膚や粘膜にできる腫瘍のうち、大きくなってまわりに広がったり、違う臓器に転移して生命に危険があるものである。1980 年以降、日本人の死因の第 1 位ががんであり、近年では死亡者の 30% (男

^a 東洋大学自然科学研究室：〒 112-8606 東京都文京区白山 5-28-20

Natural Science Lab., Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606 JAPAN

^b 東洋大学現代社会総合研究所：〒 112-8606 東京都文京区白山 5-28-20

Institute of Social Sciences, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606 JAPAN

34%、女 26%) を占めているが、世界全体ではがんは感染症、心疾患に次ぐ 3 番目の疾病であり、病気による死亡者全体の 13% である。しかし、アジア (特に中国)、アフリカ、中南米の諸国がその内の 70% を占めており、今後は途上国の生活習慣の欧米化により、世界のがんによる死亡者数は 2030 年には倍増すると予測されている。そのため、世界各国でがん撲滅の取り組みが行われており、日本では 2006 年 6 月のがん対策基本法が成立している。

人間の体は約 37 兆個の細胞から成るが、ほとんどのがんは、1 個の異常な細胞 (がん細胞) が増殖してできる。がん細胞は、正常な細胞が何らかの原因で遺伝子に傷がついたり、正常な遺伝子の突然変異によってがん細胞に変わることがある。遺伝子に突然変異を与える原因は、様々であり次のようなものが挙げられている。

- ①生活習慣：喫煙 (能動, 受動), 飲酒, 塩分・脂肪の過多, 野菜・果物・食物繊維の不足, 肥満・運動不足, 性行動
- ②生物的要因：細菌 (ピロリ菌)・ウイルス (肝炎ウイルス, パピローマウイルス等)・寄生虫の感染
- ③物理的要因：放射線 (天然, 人工)・紫外線の被曝
- ④その他：医薬品, 遺伝, ストレス等

このように複雑ながんの原因を解明するために、がんの死亡率データを解析して決定要因を探索する統計学的研究が行われており (Hiatt and Breen 2008, Reid et al. 2008, Merletti et al. 2011), これまで 3 つの手法がある。その第一は疫学的研究であり、長期間にわたる多数の被験者と対照者でのがん死亡率の差を統計的に検定し、決定要因を探るものである。Doll and Peto (1981) はこの方法で米国人のがん死亡率に対する各種の要因が占める割合を推定し、食生活要因による死亡割合が 35%, 喫煙の割合が 30%, ウイルスや細菌の感染が 10%, 性行為が 7%, 飲酒等の要因が 4% 以下と発表した。Harvard Center for Cancer Prevention (1996) も米国人について同様の推計を行い、喫煙が 30%, 食生活・肥満が 30%, 運動不足, 職業, 家族歴, ウイルス感染等の要因が各 5%, 飲酒等の要因が 3% 以下と報告した。また, Inoue et al. (2012) は日本人の 2005 年のがん死亡率を解析し、喫煙が 24%, 感染が 22%, 飲酒が 6%, 塩分摂取, 肥満, 野菜不足等が 1% 以下であるとしている。しかし、これらの結果は 1 国のデータに基づく推計値であり、対象範囲の点で限定的である。また、この方法でがんの原因を解明するためには莫大な費用と長期の時間が必要である。

第二の方法は、がんの死亡原因と考えられる各種の要因について個別にがん死亡率との相関を分析するものであり、これを用いて各種要因の寄与の大きさを論じた論文は多い。例えば, Colli (2005) は世界 71 カ国の前立腺がんの死亡率と食習慣および日光線量との相関を分析している。また, 坂巻 (2001) は世界 21 カ国の部位別がんの罹患率・死亡率と食事を含む環境要因との相関を分析している。しかし、がん死亡率との個別的な相関の大きさから各要因の寄与の大きさを推定することには問題があり、これらの方法では多数の要因の中でどれが相対的に重要であるかを決定することは難しい。

第三の方法は、がん死亡率を目的変数、幾つかの要因を説明変数として重回帰分析を行い、各種要因の寄与率を定量的に求めるものである。この方法を用いて 1 国のがん死亡原

因を分析した論文は多い。例えば, Shi et al. (2005) は米国の各州のがん死亡率について幾つかの社会経済指標を説明変数として重回帰分析を行い, 所得格差の指標であるジニ係数の寄与が最大であると報告している。巖ら (1988) は日本の肝臓がん死亡率について食品因子を説明変数とする重回帰分析を行い, アルコールと緑茶の消費量が決定要因であると報告している。また, 多尾ら (2002) は日本人女性の乳がん死亡率に対して動物性脂肪, 動物性蛋白質, 植物性脂肪を説明変数とする重回帰分析を行い, これらの3変数が乳がん死亡率の決定要因であることを示している。しかし, これらの結果は対象範囲の点でやはり限定的である。

がんの死亡原因に関してより一般的な結果を得るためには, 複数の国のがん死亡率を一括して重回帰分析することが必要である。しかし, Table 1 に示すように, これまで複数国を一括解析した研究は少数であり, また OECD 加盟国, 被援助国, あるいは地域 (例えばアフリカのみ) のように対象を限定した研究が多い。Singh et al. (2012) は最多の 184 カ国を解析しているが, 説明変数が2種と少数で説明能力に欠けるため, 得られたモデルの決定係数がきわめて低い。一方, Drain et al. (2002) は最多の 18 種の説明変数を用いて高い決定係数を得ているが, 解析したのは 82 の途上国にすぎない。このように先行研究では, 解析対象や説明変数が限定的であり, それらの結果は一般性に乏しい。がんの死亡原因に関して一般的な結果を得るために, 多数の国について多数の候補説明変数を用いて重回帰分析を行い, 決定要因を探索した論文は見当たらない。

Table 1. Literatures on statistical determinant analysis of cancer mortality

Number of countries	Number of explanatory variables	Coefficient of determination	Authors
184	2	0.504	Singh et al. (2012)
87	14	0.77	Grant (2014)
82	18	0.9	Drain et al. (2002)
50	11	0.862	Ganmaa & Sato (2005)
42	6	0.9	Hebert et al. (1998)
30	9	0.657	Stare & Jozefowicz (2008)

また, 先行研究では解析手法として OLS (ordinary least squares) 等の線形回帰分析を用いているが, 各種の要因とがん死亡率との関係は一般に線形ではなく, 複雑な相関関係を示す指標が多い。このような複雑な相関関係に対して一つの対処策と考えられるのが人工ニューラルネットワーク (ANN) である。しかし, ANN は解析能力は高いが, 過学習や局所解等, 多くの問題があるため, 最適なモデルの構築が難しいことが指摘されている。一方, 近年注目されているサポートベクターマシン (SVM) (Cristianini & Shaw-Taylor 2000, 小野田 2007, 阿部 2011) は ANN と同様の非線形解析手法であるが, ANN と比較して飛躍的な高速処理が可能であり, また, 局所解の問題がない。そのため, 現在知られている多くの手法の中で最も認識性能に優れた学習モデルの一つであると注目されている。しかし, これまで SVM をがん死亡率の決定要因解析に適用した研究は見当たらない。

そこで本研究では、がんの死亡原因について一般的な結果を得るために、世界中の 147 カ国のがん死亡率のデータを目的変数とし、食料、健康、教育、地理、環境、経済の 6 分野の 40 種の指標を候補説明変数として用いて SVM により解析し、変数最適化により決定要因を探索する実証分析を行った。筆者の知る限り、がん死亡率の決定要因について本研究のような大規模データを用いて精密な解析を行った先行研究は見当たらない。

2. 方法

2.1 がん死亡率および各種指標のデータ

がん死亡率は IARC の GLOBOCAN 2012 から人口 10 万人当たりの男女合計年齢調整全部位がん死亡率のデータを入手した。この中には、がん部位別の死亡率データが含まれるが、本研究では第 1 報として各国の全がん死亡率が世界共通の説明変数でどの程度まで再現できるかを解析した。本データには、全がん死亡率は 200 以上の国について入手可能であるが、説明変数の内の幾つかが欠損する国があるため、147 カ国を解析の対象とした。その内、死亡率上位 10 カ国および主要 10 カ国の死亡率を Table 2 に示す。死亡率最高位のモンゴル (CMR=161.0) と最低位のカーボベルデ (CMR=50.2) とは 3 倍以上の違いがある。日本のがん死亡率 (CMR=93.8) は 147 カ国の平均に近い。

Table 2. Cancer mortality rates (CMR) of highest 10 countries and main 10 countries

Country	CMR	Rank	Country	CMR	Rank
Mongolia	161.0	1	Russia	122.5	19
Dominica	155.4	2	China	122.2	20
Hungary	152.1	3	United Kingdom	110.0	36
Armenia	150.3	4	France	107.9	42
Uruguay	144.8	5	United States	105.0	43
Zimbabwe	142.7	6	Germany	100.8	57
Kazakhstan	140.2	7	Japan	93.8	76
Kenya	135.3	8	Finland	86.1	93
Uganda	134.2	9	India	64.5	135
Poland	131.0	10	Cape Verde	50.2	147

147 カ国の内訳は Table 3 に示すように、先進国、途上国（先進国・途上国の定義は World Bank による）の双方が含まれ、また、世界の全地域の国々が網羅されている。本研究では世界共通のがん死亡原因としてできるだけ一般的な結果を得るために、147 カ国のがん死亡率を一括して解析することを試みた。

Table 3. Distribution of the type of 147 countries analyzed

	Europe	Asia	America	Oceania	Africa	Total
Developed countries	25	6	3	2	0	36
Developing countries	9	29	26	2	45	111

説明変数（要因）は、WHO 等の報告や多くの論文で検証されていること、世界中の多数の国についてデータが入手可能であること等の選択理由から、Table 4 に示す 40 種の候補指標を採用した。以上の指標の内、人口当たりの数値が算出できるものはその数値を用いた。これらの指標の記述統計量が示すように、指標の中には分布の偏りがかなり大きいものがあるので、各指標は対数を用いて最小値 0 と最大値 1 の間にできるだけ均一に分布するよう正規化して解析に用いた。なお、説明変数が 40 種もあるため、変数間の相互相関係数は紙面の都合上、割愛せざるを得ない。ただし、SVM による解析では OLS と異なり、変数間に強い相関がある場合でも解析可能であり、多重共線性問題は生じない。

2.2 サポートベクターマシンによる解析

SVM のソフトウェアは LIBSVM ver.2.89 (Chang & Lin) の回帰機能 (ε SVR)、カーネル関数は RBF を用いた。候補説明変数の中から決定要因を探索するためには SVM の 3 種のモデルパラメータ、 g (RBF カーネルの gamma)、 c (cost)、 p (loss function の ε)、および説明変数の最適化が必要である。本研究では、モデルパラメータの最適化には交差検証法 (CVT) を採用し、説明変数の最適化には感度分析法を採用した。感度分析法は、各指標の感度を計算し、その感度の低い指標を順次削除しながら SVM 解析を行って予測値と実測値の平均二乗誤差 (RMSE) が最小となる点を探索する方法であり、筆者らは様々な問題に適用し、その有効性を検証している (Tanabe et al. 2013, 田辺・鈴木 2013, 2014)。そこでこれら 2 種の最適化を組み合わせた以下の手順を用いて決定要因の探索を行った。

- ①全データを 10 群に分割し、第 1 群を予測セット、その他の群全体を学習セットとする。
- ②学習セットについて 3 つのパラメータ g , c , p をグリッドサーチして RMSE の最小点を探し、このモデルに予測セットの指標値を入力してがん死亡率の予測値を求める。
- ③第 2 群以下の各群を予測セットとして以上の操作を繰り返し、全データの RMSE を求める。
- ④各指標の感度を求めるために、当該指標は実際の数値に設定し、その他の指標は平均値に設定したデータを予測セットとしてモデルに入力し、出力値を求める。
- ⑤当該指標の設定値を説明変数、出力値を目的変数とする単回帰分析を行い、回帰直線の傾きをその指標の感度とする。
- ⑥全指標の中で感度の絶対値の最も小さい指標を取り除き、以上の操作を繰り返す。
- ⑦指標数とパラメータ g , c , p の組み合わせの中で、全データの RMSE が最小になる指標の組み合わせを決定要因とする。

Table 4. Definition, units, data source, and descriptive statistics of dependent and explanatory variables

Variable	Definition and unit	Field	Source	max	min	average	skewness
Cancer mortality rate	Mortality rate of all cancers per 100,000 population		IARC	161.00	50.20	95.52	0.362
Grain self-sufficiency rate	Proportion of domestic grains among those consumed in the country	Food	FAO	286.9	0.0	72.7	1.055
Meat consumption	Meat consumption: per capita in kg per year	Food	FAO	133.1	3.5	45.8	0.624
Fish consumption	Fish consumption per capita in kg	Food	FAO	91.00	0.00	16.29	1.729
Milk consumption	Milk consumption per capita in kg per capita per year	Food	FAO	361.2	3.5	111.0	0.737
Vegetables consumption	Vegetables consumption in gram per capita per day	Food	FAO	910.0	21.0	241.1	1.351
Fat intake	Average daily fat intake per capita in gram	Food	FAO	164.0	20.0	82.1	0.653
Sodium intake	Sodium intake per capita in gram per day	Food	Powles	5.98	1.48	3.48	0.250
Sugar intake	Sodium intake per capita in gram per day	Food	FAO	166.0	12.0	84.3	-0.075
Coffee consumption	Coffee consumption per capita in kg per year	Food	WHO	12.00	0.00	1.87	1.744
Smoking prevalence	Percentage of men and women ages 15 and over who smoke daily	Health	WHO	41.82	6.05	21.51	0.237
Alcohol consumption	Alcohol consumption ages 15 and over in litres per capita per year	Health	WHO	14.37	0.00	5.12	0.283
Obesity	Rate of adults of body mass index over 30	Health	WHO	35.01	1.00	15.34	0.187
Calorie intake	Daily calorie intake per capita in 10 ³ kcal	Health	FAO	3825	2061	2901	0.096
Undernourishment	Prevalence rate of undernourishment	Health	WHO	50.74	3.86	12.95	1.243
Physicians	Number of physicians per 1,000 population	Health	WHO	6.51	0.01	1.62	0.737
Hospital beds	Number of hospital beds per 1,000 population	Health	WHO	13.71	0.23	3.05	1.343
Health expenditure	Sum of public and private health expenditures per capita	Health	WHO	17.07	2.00	6.74	0.828
HBV	Hepatitis B virus infection rates	Health	WHO	19.00	0.30	6.22	0.358
HCV	Hepatitis C virus infection rates	Health	WHO	14.00	0.20	2.35	2.760
HPV	Human papilloma virus incidence rate	Health	WHO	46.50	1.60	17.89	0.529

Table 4. (Continued)

Life expectancy	Expected number of years of life remaining at birth	Health	WHO	83.0	48.3	69.6	-0.713
Senior rate	Population rate of ages 65 and above	Health	WB	25.20	0.67	8.30	0.874
School life expectancy	School life expectancy, primary to tertiary education	Education	UN	18.25	4.14	11.82	-0.407
Literacy rate	Literacy rate of people ages 15 and above	Education	UN	100.0	28.7	84.5	-1.322
Intelligence quotient	Ratio of tested mental age to chronological age	Education	Eutimes	106.0	63.6	84.8	-0.208
Education expenditure	Public education expenditure per capita in \$	Education	UN	13.12	0.98	4.61	1.161
Religion	Rate of population who say religion is important	Education	Gallup	99.5	0.1	55.4	-0.438
Latitude	Average latitude of country	Geography	WB	64.17	0.25	26.79	0.317
Temperature	Annual average temperature	Geography	WB	29.50	0.10	18.51	-0.627
Precipitation	Average precipitation depth in mm per year	Geography	WB	3028	51	1135	0.621
Forest area	Forest areas per capita in m ²	Environment	FAO	94.22	0.07	30.15	0.522
Terrestrial protected area	Rate of terrestrial protected areas to total land area	Environment	UN	53.60	0.02	13.52	0.997
Water access	Rate of population with access to improved water source	Environment	WHO	100.0	47.9	87.4	-1.246
Sanitation	Rate of population using improved sanitation facilities	Environment	WHO	100.0	9.0	72.7	-0.812
Air pollution	Country level of PM10 in $\mu\text{g}/\text{m}^3$	Environment	WHO	161.9	7.2	48.6	1.449
Urban population	Rate of people living in urban areas	Environment	UN	98.2	13.9	58.0	-0.234
Gross domestic product	Per capita gross domestic product in 10 ³ \$	Economy	IMF	80595	553	14605	1.425
Poverty rate	Poverty headcount ratio at national poverty line per population	Economy	UN	80.0	0.0	28.5	0.759
HDI	Human Development Index	Economy	UN	0.955	0.304	0.675	-0.359
Senior worker	Rate of economically active population ages 65 and over	Economy	ILO	91.24	1.64	26.99	0.819

For more detailed definition of variables, see homepages of data sources.

3. 結果と考察

40種の候補指標から出発して低感度の指標を逐次削除しつつ決定要因を探索した。ただし、147カ国の内にはがん死亡率の予測誤差が全体のRMSEの3倍を超えるものがあり、これらを含めた147カ国のまま変数探索を行うと、全体のRMSEがかなり大きくなり、回帰決定係数が低下した。そこで、このような予測誤差が全体のRMSEの3倍を超える5カ国はOutliers（外れ値）として学習データから除外して決定要因探索を行ったところ、指標14種で予測値と実測値とのRMSEが妥当な最小値となった。したがって、この14種の指標が世界142カ国のがん死亡率の決定要因となる。

Fig. 1に示すように、Outliers5カ国を除く142カ国全体でのがん死亡率の実測値と予測値とのRMSEは2.39、回帰決定係数（ R^2 ）は0.730となり、世界の多数の国のがん死亡率

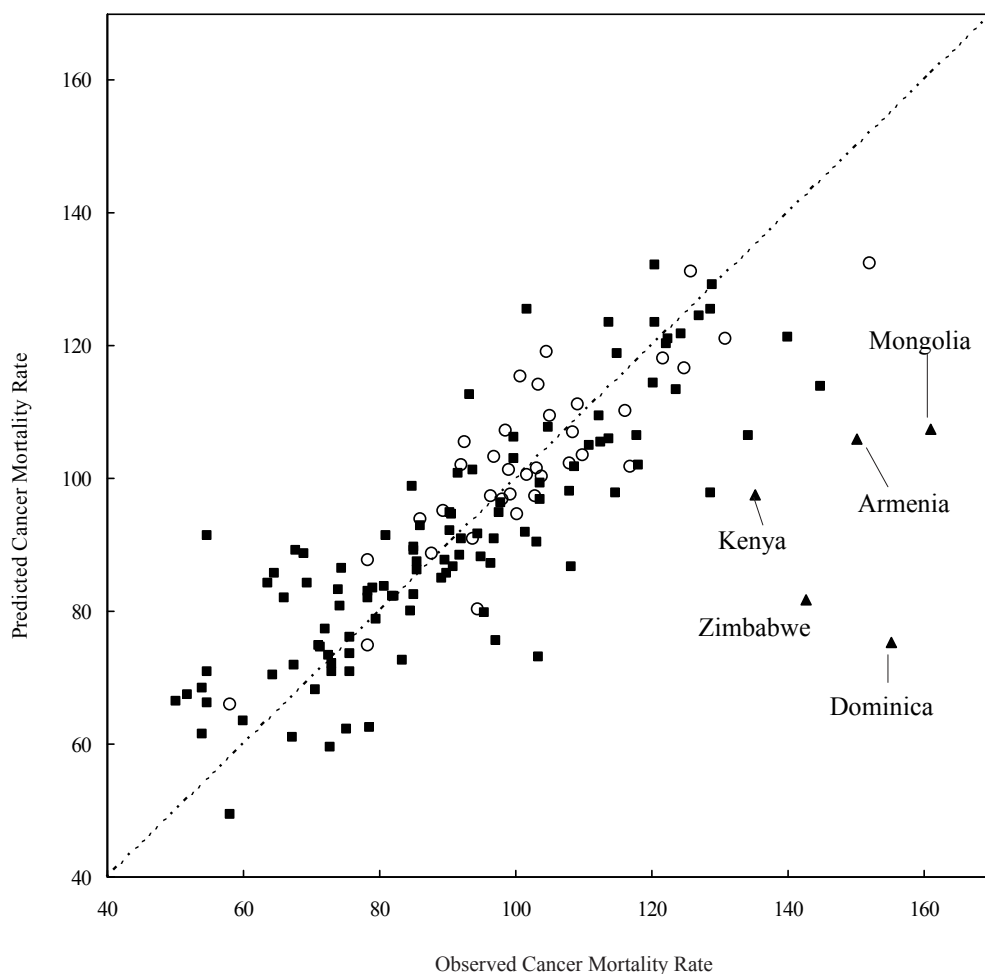


Fig. 1. Scatter plot of observed versus predicted cancer mortality rates of developed countries (open circle), developing countries (filled square), and outlying countries (filled triangle)

が先行研究よりはるかに高い精度で再現できることを確認した。ただし、先進国では予測値と実測値の一致が概ね良いが、途上国では誤差が先進国より大きい。これには途上国についての各種指標データの信頼性に問題がある可能性が考えられる。

決定要因 14 種を含む 40 種の説明変数について感度分析法で得られた感度と、がん死亡率との相関係数との散布図を Fig. 2 に示す。相関係数と感度とは相関が低く、特に、決定要因の中には相関係数がかなり低いにもかかわらず感度が高いものが多数ある。このことから、先行研究では、

- ①がん死亡率との相関の高さに基づいて指標の寄与率を分析している論文が多いが、このような議論の結果には疑問がある、
 - ②説明変数を選択する際、がん死亡率との相関係数に基づいて選定している論文が多いが、この方法では死亡率に大きく寄与している要因を見逃す可能性がある、
- と言える。それに対して、本研究では、線形、非線形を問わず、あらゆる相関関係の解析が可能な SVM を用いて解析し、多種多様な分野の多数の候補変数の中から変数選択により決定要因を探索し、予測精度のよいモデルが得られ、がん死亡率の決定要因に関して既往の研究より信頼性の高い結果を得ることができたと考えられる。

決定要因の探索過程においてモデル学習から外した 5 カ国について誤差がきわめて大きくなった原因を考察すると、次の 2 つの可能性が考えられる。第一は本研究で予測に用い

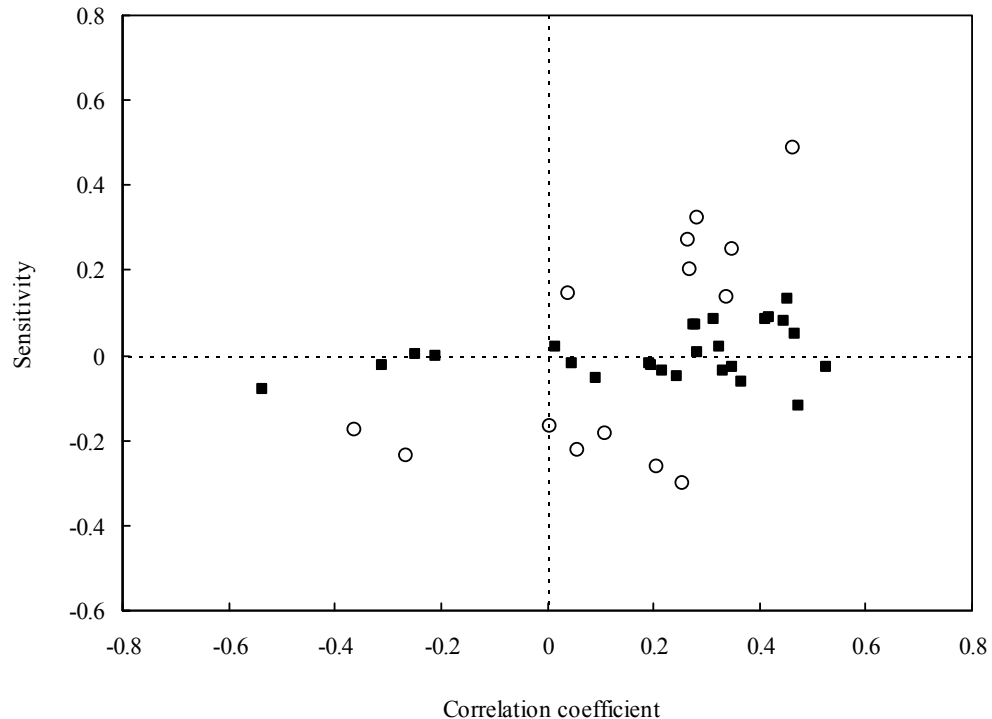


Fig. 2. Scatter plot of correlation coefficients versus sensitivities of 14 determinants (open circle) and other explanatory variables (filled square)

た方法に付随する問題である。これら Outliers 5 カ国は Fig. 1 に示すようにいずれも死亡率の実測値が高い。このことから、本研究で SVM モデルの学習・予測に用いた交差検証法 (CVT) では、がん死亡率が最大付近の国については、それ以外の国のデータで学習したモデルを用いて外挿で予測することになるため、内挿で予測する内側の国より予測誤差が大きくなる可能性があることが考えられる。

第二の原因としては、本研究において作成した世界 142 カ国の死亡率を統一的に予測するモデルで決定要因となった以外の要因が、Outliers 5 カ国のがん死亡率に特に大きく働いている可能性が考えられる。そこで、147 カ国の中でがんの死亡率が最大であるモンゴルについて予測誤差の原因を解析すると、同国は他国と異なるがん死亡率の特徴があることが分る。すなわち、モンゴルの死亡率は肝がんと胃がんは世界 1 位と高く、食道がんも 4 位と高いが、世界 1 位のがんは 57 位、乳がんは 147 位、大腸がんは 106 位であり、世界全体の平均順位と大きく異なっている。同国で肝がん死亡率が高い原因として、B 型肝炎ウイルス (HBV) および C 型肝炎ウイルス (HCV) の感染率が他国より突出して高いためと報告されているが、本研究で用いたデータでも HBV と HCV 感染率の世界順位は 6 位と 3 位と高い。一方、胃がん死亡率については、その最大原因と考えられている塩分摂取量が 8 位と高く、しかも 1 位のカザフスタンとは摂取量の差が非常に小さい。そのため胃がん死亡率が 1 位、食道がん死亡率が 4 位になっていると考えられる。これに対して、タバコ喫煙率は 41 位と高くないため、肺がん死亡率が低く、また、脂肪摂取量の順位も 68 位と低いため、大腸がんも 106 位と低い。このように、モンゴルではがん死亡率順位の世界平均が高い肺がん、乳がん、大腸がんの割合が低く、逆に世界順位の低い胃がんと食道がんの順位が高いという特異的事情が、がん死亡率世界予測モデルで同国の予測誤差を著しく大きくさせたと考えられる。

モンゴル以外の Outliers 4 カ国 (ドミニカ、アルメニア、ジンバブエ、ケニア) についてもモンゴルと同様の特殊要因が関与していると考えられる。たとえば、ドミニカは男性の前立腺がんの死亡率が 3 位、ジンバブエは女性の子宮頸がんが 4 位、食道がんが 12 位、前立腺がんが 17 位、アルメニアは膀胱がんが 1 位、卵巣がんが 4 位、ケニアは食道がんが 2 位と、いずれも世界平均順位が低いがんの死亡率がきわめて高いという特異性が見られる。したがって、Outliers 5 カ国のがん死亡率の予測誤差が他の 142 カ国の RMSE の 3 倍以上になった原因はこれらの国の食生活を含む生活習慣の特異性にあると考えられる。

決定要因 14 種の内訳、その感度、および感度 S_i から次式 (1) により計算した死亡率に対する寄与率 CR_i を Table 5 に、また、寄与率を分野別に集計した結果を Table 6 に示す。

$$CR_i(\%) = \frac{S_i^2}{\sum_{i=1}^{14} S_i^2} \times 100 \quad (1)$$

第 1 は第 1 位の知能指数の感度の符号が正で、かつ寄与率がきわめて高いことである。常識的には知能指数の低い途上国ほど食品、医療、衛生 (上下水道) 等の生活上の問題が多く、がん死亡率が高くなるのではないかと予想される。しかし、知能指数が高いほど、

Table 5. Sensitivities (S), contribution rates (CR) and correlation coefficients (CC) of 14 determinants of cancer mortality rates

Determinant	Field	S	CR (%)	CC
Intelligence quotient	Education	0.4894	26.1	0.4624
Vegetables consumption	Food	0.3235	11.4	0.2827
Life expectancy	Health	-0.3025	10.0	0.2534
Health expenditure	Health	0.2694	7.9	0.2646
Sodium intake	Food	-0.2613	7.4	0.2081
Meat consumption	Food	0.2498	6.8	0.3503
Poverty rate	Economy	-0.2363	6.1	-0.2659
Education expenditure	Education	-0.2217	5.4	0.0582
HPV	Health	0.2037	4.5	0.2689
Urban population	Environment	-0.1840	3.7	0.1109
Senior worker	Economy	-0.1740	3.3	-0.3621
Fish consumption	Food	-0.1686	3.1	0.0068
Precipitation	Geography	0.1453	2.3	0.0384
Grain self-sufficiency rate	Food	0.1387	2.1	0.3399

Table 6. Total contribution rates of determinants of six fields

Field	Number of determinants	Total contribution rate (%)
Education	2	31.4
Food	5	30.8
Health	3	22.4
Economy	2	9.4
Environment	1	3.7
Geography	1	2.3

野菜や肉の摂取量が多いほど、医療費が多いほど、貧困率が低いほどがん死亡率が高いという本研究の結果は、先進国ほど死亡率が高く、がんが文明病であることを示唆していると解釈できる。先行研究の中では、Singh et al. (2012) は 184 カ国の子宮がんの死亡率を解析し、人間開発指数 (Human Development Index: HDI) と貧困率が重要な決定要因になると報告している。しかし、HDI が大きい国ほど、また貧困率が低い国ほど、すなわち先進国ほど死亡率が低くなるとしており、本研究の結果とは逆である。この違いの原因は本研究では全がんの死亡率を解析したのに対し、彼らは子宮がんの死亡率を解析したが、子宮がんの原因であるパピローマウイルス (HPV) の感染率はその国の先進度と高い関連性があるためである。本研究の結果でも HPV は決定要因に入り、その符号が正であることから、彼らの結果と合致していると言える (しかし、彼らは HPV を説明変数に取り上げていない)。

第 2 は食品指標が決定要因の中に 5 指標も入り、それらの寄与率の合計が約 3 割を占めている点である。特に野菜摂取量は高い寄与率を示している。しかし、本研究ではこれ

ら5種の指標以外に、牛乳、砂糖、コーヒーの3指標も候補説明変数に取り上げたが、これらの指標はどれも決定要因にはならなかった。これら食品指標については先行研究では Ganmaa & Sato (2005) は50カ国の乳・卵巣・子宮体がんの死亡率に対して各種の食品を説明変数として解析し、肉、牛乳、チーズを決定要因としている。Hebert et al. (1998) は42カ国の前立腺がん死亡率を解析し、穀物、肉、豆等が決定要因になるとしている。Stare & Jozefowicz (2008) はOECD30カ国の全がんおよび部位別がん死亡率についてOLS解析を行い、脂肪、果物等を決定要因としている。しかし、これらの先行研究ではいずれも説明変数の範囲、および解析対象国が限定的であり、得られた結果の一般性には疑問が残る。

第3は健康指標が決定要因の中に3種入り、特に平均寿命と医療費の寄与率が高いという結果である。本研究ではこれら3種の指標以外に、タバコ、飲酒、肥満、カロリー、栄養失調、医者、病床、HBV、HCV、高齢化率の10指標も候補説明変数に取り上げたが、これらの指標はどれも決定要因にはならなかった。特に、タバコ、飲酒、肥満は前記の Doll and Peto (1981), Harvard Center for Cancer Prevention (1996), Inoue et al. (2012) においてがん死亡率に対して寄与が高いとされている要因であるが、本研究の結果はこれらと明らかに異なっている。また、Grant (2014) は87カ国の全がんおよび部位別がん死亡率についてOLS解析を行い、カロリーと飲酒が決定要因になるとしている。Hebert et al. (1998) も Stare & Jozefowicz (2008) もアルコールが決定要因に含まれることを報告している。アルコールの代謝分解物であるアセトアルデヒドには発がん性が認められており、食道がんなどの危険因子にアルコールが挙げられている（鈴木, 2011）。

最後に、経済指標は決定要因の中に2種しか入らず、その寄与率の合計も1割に満たず、がん死亡率に対する寄与率は非常に低いように見える。しかし、国民所得等の経済指標は直接的な寄与が低いが、他の食品、健康、インフラ、文化等の向上に大きな寄与を与えることは明らかである。すなわち、GpC に対して学歴（相関係数0.721）、下水道（0.624）、医師数（0.610）等、相関の高い指標が存在し、これらはがん死亡率に間接的に寄与する。したがって、経済的要因はがん死亡率に対して直接的効果は低いが、間接効果は大きいといえるであろう。

4. 結論

世界147カ国の男女合計年齢調整全部位がん死亡率を目的変数、食品、健康、教育、地理、環境、経済の6分野の40種の指標を説明変数として用い、SVMモデルを学習し、感度分析法により指標を最適化した。その結果、モンゴルなどの途上国5カ国を除いた142カ国について、14種の指標でがん死亡率が回帰決定係数（ R^2 ）0.730という先行研究より高い精度で再現できること、および、14種の指標の中では教育費等の教育要因、野菜摂取量等の食品要因、および医療費等の健康要因の寄与率が全体の過半を占めることが明らかになった。

一方、本研究ではGDP等の経済的指標はがん死亡率に対する感度が低く、いずれも決

定要因とはならなかったが、経済指標は医療費、教育費等を通じてがん死亡率低減に間接的に寄与していることは自明である、すなわち、寿命や死亡率等の健康目標に対する社会的決定要因については多数の原因と結果が相互に複雑な階層構造を形成すると考えられている（近藤 2005, 川上ほか 2006, 近藤 2010）。このような階層構造問題に対するアプローチとしては、共分散構造分析手法を用いた因果構造モデルの解析がある（豊田ほか 1992）。このような因果構造を考慮したがん死亡率の決定要因の分析は本研究の結果の展開として重要なテーマである。また、世界中のあらゆる国のがん死亡率を本モデルよりも高い精度で再現するモデルを構築するためには、今後、部位別のがん死亡率データを解析することでも必要であろう。

参考文献

- GLOBOCAN 2012, "Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012," http://globocan.iarc.fr/old/summary_table_site-html.asp?selection=290&title=All+cancers+excl.+non-melanoma+skin+cancer&sex=0&type=1&window=1&africa=1&america=2&asia=3&europe=4&oceania=5&build=6&sort=0&submit=%A0Execute%A0.
- Chang, C. C., and Lin, C. J. "LIBSVM: A Library for Support Vector Machines," <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, accessed 2011/12/1.
- Cristianini N, and Shaw-Taylor J. "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," Cambridge University Press (2000); 大北剛（訳）『サポートベクターマシン入門』共立出版（2005）.
- Doll, R., and Peto, R. "The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today," *J. Natl. Cancer Inst.* **66**, pp. 1192-1309 (1981).
- Drain, P. K., Holmes, K. K., Hughes, J. P., and Koutsky, L. A. "Determinants of Cervical Cancer Rates in Developing Countries," *International Journal of Cancer*, **100** (2) pp. 199-205 (2002).
- Eutimes, The European Union Times, <http://www.eutimes.net/2009/11/iq-by-country/>, accessed 2013/2/1.
- FAO, United Nations Food and Agriculture Organization Statistics Division, <http://faostat.fao.org/site/567/default.aspx#ancor>, accessed 2013/2/1.
- Gallup, Gallup Global Reports, <http://www.gallup.com/poll/142727/religiosity-highest-world-poorest-nations.aspx#2>, accessed 2013/2/1.
- Ganmaa, D., and Sato, A. "The Possible Role of Female Sex Hormones in Milk from Pregnant Cows in the Development of Breast, Ovarian and Corpus Uteri Cancers," *Medical Hypotheses* **65** (6) pp. 1028-1037 (2005).
- Grant, W. B. "A Multicountry Ecological Study of Cancer Incidence Rates in 2008 with Respect to Various Risk-Modifying Factors," *Nutrients* **6**, pp. 163-189 (2014); doi:10.3390/nu6010163.
- Harvard Center for Cancer Prevention "Harvard Report on Cancer Prevention. Volume 1:

- Causes of Human Cancer," *Cancer Causes Control*, **7**, S3-S59 (1996).
- Hebert, J. R., Hurley, T. G., Olendzki, B. C., Teas, J., Ma, Y., and Hampl, J. S. "Nutritional and Socioeconomic Factors in Relation to Prostate Cancer Mortality: a Cross-National Study," *J Natl Cancer Inst.* **90** (21) pp. 1637-1647 (1998).
- Hiatt, R. A., and Breen, N. "The Social Determinants of Cancer: A Challenge for Transdisciplinary Science," *Am J Prev Med* **35** (2S) S141-S150 (2008).
- ILO, International Labour Union LABORSTA, <http://laborstailo.org/>, accessed 2013/2/1.
- IMF, International Monetary Fund World Economic Outlook Database, <http://www.imf.org/external/pubs/ft/weo/2013/01/weodata/index.aspx>, accessed 2013/2/1.
- Inoue, M., Sawada, N., Matsuda, T., Iwasaki, M., Sasazuki, S., Shimazu, T., Shibuya, K., Tsugane, S., "Attributable Causes of Cancer in Japan in 2005 – Systematic Assessment to Estimate Current Burden of Cancer Attributable to Known Preventable Risk Factors in Japan," *Annals of Oncology* **23**, pp. 1362-1369 (2012).
- Merletti, F., Galassi, C., and Spadea, T. "The Socioeconomic Determinants of Cancer," *Environmental Health*, **10** (Suppl 1) S1-S7 (2011). <http://www.ehjournal.net/content/10/S1/S7>.
- Reid, K., Riemsma, R., and Kleijnen, J. "Preventability of Cancer by Food, Nutrition, Physical Activity and Weight Management," Systematic Literature Review commissioned by WCRF International (2008). http://www.dietandcancerreport.org/docs/Preventability_Review%20of%20previous%20estimates.pdf.
- Shannon, M. S., and Jozefowicz, J. J. "The Effects of Environmental Factors on Cancer Prevalence Rates and Specific Cancer Mortality Rates in a Sample of OECD Developed Countries," *International Journal of Applied Economics*, **5** (2) pp. 92-115 (2008).
- Shi, L., Macinko, J., Starfield, B., Politzer, R., Wulu, J., Xu, J. "Primary Care, Social Inequalities and All-cause, Heart Disease and Cancer Mortality in US Counties: A Comparison between Urban and Non-urban Areas," *Public Health*, **119**, pp. 699-710 (2005).
- Singh, G. K., Azuine, R. E., and Siahpush, M. "Global Inequalities in Cervical Cancer Incidence and Mortality are Linked to Deprivation, Low Socioeconomic Status, and Human Development," *International Journal of MCH and AIDS*, **1** (1) pp. 17-30 (2012).
- Tanabe, K., Kurita, T., Nishida, K., Lucic, B., Amic, D., Suzuki, T. "Improvement of carcinogenicity prediction performances based on sensitivity analysis in variable selection of SVM models," *SAR QSAR Environ Res* **24**, pp. 565-580 (2013).
- UN, United Nations Development Program Human Development Report, <http://hdr.undp.org/en/>, accessed 2013/2/1.
- WB, World Bank World Development Indicators, <http://data.worldbank.org/indicator>; World Bank Global Development Network Growth Database, <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,contentMDK:20701055~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html>, accessed 2013/2/1.
- WHO, World Health Organization World Health Statistics, <http://www.who.int/gho/>

- publications/world_health_statistics/en/index.html, accessed 2013/2/1.
- 阿部重夫『パターン認識のためのサポートベクトルマシン入門』森北出版 (2011).
- 小野田崇『サポートベクターマシン』オーム社 (2007).
- 川上憲人・小林廉毅・橋本英樹 (編)『社会格差と健康—社会疫学からのアプローチ』東京大学出版会 (2006).
- 近藤克則『健康格差社会—何が心と健康を蝕むのか』医学書院 (2005).
- 近藤克則『幸福・健康の社会的決定要因—社会疫学の視点から』科学 (80) pp. 290-294 (2010).
- 坂巻路可, 廣畑富雄『世界各国におけるがんの罹患率・死亡率と食事を含む環境要因との関連について』中村学園研究紀要, (33) pp. 191-202 (2001).
- 鈴木孝弘『生命と健康百科』駿河台出版社 (2011).
- 多尾清子・吉井健悟・田中章太郎・大井達雄・有田清三郎『乳癌死亡率に関わる要因の統計分析 - わが国における既存統計から - 』関西医科大学教養部紀要, **22**, pp. 24-33 (2002).
- 田辺和俊・栗田多喜夫・西田健次・鈴木孝弘『サポートベクター回帰を用いた 158 カ国の国債格付けの再現』情報知識学会誌, **23** (1) pp. 70-91 (2013).
- 田辺和俊・鈴木孝弘『サポートベクターマシンを用いたエコロジカル・フットプリント値の決定要因の分析』日本エネルギー学会誌 **92**, pp. 1207-1213 (2013).
- 田辺和俊・鈴木孝弘『サポートベクターマシンを用いた世界各国の幸福度の決定要因の実証分析』経済分析 (188) pp. 44-67 (2014).
- 田辺和俊・鈴木孝弘『サポートベクターマシンを用いた世界各国の平均寿命の決定要因の実証分析』厚生指標, **61** (13), pp.23-30 (2014).
- 豊田秀樹・前田忠彦・柳井晴夫『原因を探る統計学 共分散構造分析入門』講談社 (1992).
- 嚴善紹・林恭平・渡辺能行・小笹晃太郎・東あかね・青池晟・川井啓市・松永栄『肝臓癌死亡とその関連因子についての地理疫学的研究』日本衛生学雑誌, **43** (5), pp. 987-994 (1988).