一般口演 | 医療データ解析

一般口演11
機械学習

2021年11月20日(土) 09:10 〜 11:10 G会場 (2号館3階232+233)

[3-G-1-07] 原死因決定プロセスの効率化に資する機械学習による原死因コード変更予測

*大井川 仁美[1]、今井 健[1]、香川 璃奈[2]、明神 大也[3]、今村 知明[3] （1. 東京大学大学院医学系研究科疾患生命工学センター, 2. 筑波大学医学医療系, 3. 奈良県立医科大学公衆衛生学講座）

*Hitomi Oigawa[1], Takeshi Imai[1], Rina Kagawa[2], Myojin Tomoya[3], Imamura Tomoaki[3] （1. 東京大学大学院医学系研究科疾患生命工学センター, 2. 筑波大学医学医療系, 3. 奈良県立医科大学公衆衛生学講座）

キーワード：Underlying Cause of Death, Automated ICD coding, Machine Learning, Artificial Intelligence

【背景と目的】日本では、原死因決定を行うために独自の人口動態死因オートコーディングシステムを利用しているが、約4割の原死因決定は自動的に行われず、人手による確定作業が行われている。その主な要因は、傷病名以外の手術欄や解剖所見、外因の発生状況など「付帯情報」の記載である。しかし、実際に付帯情報の人手確認により仮原死因コードの修正に至るケースは少なく、作業の効率化のためこれらを自動的に弁別する手法の確立が求められている。そこで本研究では機械学習ベースの分類モデルにより、付帯情報の影響による仮原死因の変更有無が判別可能かの調査を行う。

【方法】統計法33条に基づき提供を受けた H27-H30の死亡票実データからランダム抽出した50万件に対し、標準病名マスター等を用い傷病名に ICD-10コードを付与した。全ての傷病名に ICD-10コード付与できた死亡票（約6割）に対し、原死因選択をするフリーソフトウェア Irisを用いて仮原死因を決定した。また I 欄 II 欄各病名の ICD10 コード、付帯情報の各項目の有無、 Iris が付与した仮原死因を入力とし、分類器学習モデルとして汎用的な勾配ブースティング決定木の一種である XGBoostを用いて、確定原死因が仮原死因から変更されるか否かを予測するモデルを構築し精度を算出した。

【結果と考察】本手法による変更予測モデルの精度は約9割で、重要度の高い因子は、年齢、備考欄の記載の有無、 I501コードの存在、その他不言すべき事柄の記載の有無、手術年月日の有無などであった。記載の有無だけを用いたベースライン手法でも高い精度を達成しており、非常に有望な手法と考えられた。今後自然言語記載内容の分散表現の利用等でさらなる精度向上と変更後のコード提示への発展が見込まれ、人手確認によって行われてきた原死因確定作業の大幅な効率化、負荷軽減が図れると期待される。

原死因決定プロセスの効率化に資する機械学習による原死因コード変更予測

大井川　仁美[*1]、今井　健[*1]、香川　璃奈[*2]、明神　大也[*3]、今村　知明[*3]
*1 東京大学大学院医学系研究科疾患生命工学センター、*2 筑波大学医学医療系、
*3 奈良県立医科大学公衆衛生学講座

Prediction of Changes in the Underlying Cause of Death Code by Machine Learning

Hitomi Oigawa[*1], Takeshi Imai[*1], Rina Kagawa[*2], Tomoya Myojin[*3], Tomoaki Imamura[*3]
*1 Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo,
*2 Faculty of Medicine, University of Tsukuba,
*3 Department of Public Health, Health Management and Policy, Nara Medical University,

[Background and purpose] In Japan, about 40% of the underlying causes of death are determined manually. The main factor is the description of incidental information such as surgery and anatomical findings. This study investigates whether machine learning can determine whether the underlying cause of death has changed due to the influence of incidental information. [Method] The target data is 500,000 randomly selected from death votes from 2015 to 2018. We converted the name of the injury or illness on the death votes into an ICD-10 code (about 60%) and used the free software Iris to determine the temporary underlying cause of death. We used XGBoost to build a model that predicts whether or not the temporary underlying cause of death has changed, and calculated the accuracy. The input data was the ICD10 code of the name of the injury and illness, the presence or absence of each item of the incidental information, and the temporary underlying cause of death given by Iris. [Results and discussion] The accuracy of the change prediction model by this method was about 90%. Factors of high importance were age, presence or absence of remarks, presence of I501 code, etc. By developing this method, it is expected that the efficiency of manual work to determine the underlying cause of death can be greatly improved.

Keywords: Underlying Cause of Death, Automated ICD Coding, Machine Learning, Artificial Intelligence

1. 背景と目的

人口動態調査の生成物の 1 つである死因統計は、我が国そして世界の保健に関する重要な資料である。死因統計の製表では、死亡診断書から生成される人口動態調査死亡票（以下、死亡票）から抽出した原死因が用いられる。原死因とは、死亡診断書に記載される傷病名のうち「直接に死亡を引き起こした一連の事象の起因となった疾病又は損傷」または「致命傷を負わせた事故又は暴力の状況」を指し、WHO が定める原死因選択ルールに則って抽出される [1]。ただし、ルール適用が困難な場合も存在し、適切な対処が必要である。

厚生労働省では、年間約 135 万件以上の死亡票を処理するため、独自開発した人口動態死因オートコーディングシステム（以下、オートコーディングシステム）を使用している。オートコーディングシステムは、死亡票の「死亡の原因」の I 欄と II 欄（以下、I 欄 II 欄）にある傷病名に対する ICD-10 コードの付与と、原死因の仮付与・確定を行う。しかし、全死亡票がオートコーディングシステムのみで完結することはなく、傷病名に対する ICD-10 コード付与が困難である場合や仮付与した原死因（以下、仮原死因）に疑義がある場合、I 欄 II 欄以外の項目（手術欄や解剖所見、「その他特に付言すべきことがら」など。以下、付帯情報）に何かしら記載がある場合は、人手による原死因確定作業が実施される。現在、人手による作業は死亡票全体の約 4 割であり、付帯情報の確認作業が主である [2][3]。日本における年間死亡数は年々増加しており、より正確で効率のよい原死因確定が求められている。

このような背景の元、著者らは厚生労働科学研究費補助金研究事業「死因統計の精度及び効率性の向上に資する機械学習の検討に関する研究」において、機械学習を適用することで、原死因確定プロセスにおける人手の作業の効率化を目指し、付帯情報による仮原死因の変更の有無を機械学習し、変更がある場合はその内容も示すことを目標としている。これまでの調査により、付帯情報の確認作業後に仮原死因が変更されるケースは少ないことが明らかになっており、これを高精度で自動的に分類することで確認作業の効率化が可能になることが判明している。また、変更内容を提示することで人手によって詳細を確認するべき部分についても効率性・正確性の向上に資すると考えられる。

これらの手法開発の上で必要なデータは、

① 死亡票データ
② I 欄 II 欄記載の傷病名に対するオートコーディングシステムが付与した ICD-10 コード
③ 原死因選択ルールに則りオートコーディングシステムが付与した仮原死因の ICD-10 コード
④ 最終的に確定された原死因（以下、確定原死因）の ICD-10 コード

である。しかし、オートコーディングシステムは非公開のため、②および③は入手できない。②に関しては、標準病名マスター等の利用により、ある程度の対処が可能であるが、③は煩雑な原死因ルールの実装を行わなければならない。そこで、著者らはこれまでオートコーディングシステムに類似したフリーソフトウェア Iris による代替を試み [4][5]、その結果、Iris がオートコーディングシステムの代替として有効であることが示唆され、研究遂行に必要な①～④のデータを得ることが可能になっている。

そこで、本研究では、①～④のデータを使用して、仮原死因の変更有無を対象にした機械学習の効果について調査を行うことを目的とする。具体的には、機械学習ベースの分類

モデルを使用して、付帯情報の影響による仮原死因の変更有無判別を実施する。本研究では今後の手法開発におけるベースライン手法として、まず自由記載を含む付帯情報の内容は考慮せず、付帯情報の各項目の記載の有無(2 値情報)だけを用いる最も簡単なモデルを構築した。

2. 方法

統計法 33 条に基づき提供を受けた平成 27 年度から平成 30 年度の死亡票データ約 520 万件から 50 万件の死亡票データをランダム抽出した。そして、標準病名マスター[6]等を用い各死亡票の傷病名に ICD-10 コードの付与を行い、全ての傷病名に ICD-10 コード付与できた死亡票 320,112 件(320,112/500,000 = 64.0%)に対し、Iris を用いて仮原死因を決定した。最終的に仮原死因が決定でき、確定原死因と比較可能であるデータは 320,008 件となり、その一致率は 91.0%であった。また、仮原死因の変更有無と付帯情報有無をまとめたものを表 1 に示す。このデータを本研究対象の死亡票データとし、以下の手順で機械学習に用いる。

2.1 機械学習用データの作成

死亡票データと 320,008 件で使用された仮原死因の ICD-10 コードのデータ(1553 種類)を組み合わせて、機械学習用のデータを生成した。仕様は表 2 の通りである。

学習に用いた入力データは 1577 次元のベクトルであり、1 列目に年齢、2 列目に性別が続く。3 列目から 1555 列目までは、死亡票に出現する仮原死因の ICD-10 コードである。対象の死亡票の傷病名に該当する ICD-10 コードがあり、それが仮原死因である場合は無条件に 1 を付与し、それ以外にⅠ欄Ⅱ欄に記載がある場合は、原死因としての選択されやすさを鑑み、Ⅰ欄のエ欄、ウ欄、イ欄、ア欄そしてⅡ欄という優先順位に従って、0.85 から順に 0.70, 0.55, 0.40, 0.25 と重みをつけた。例えば、I 欄のア、イにのみ病名が記載され、イ欄病名が仮原死因として選択された場合、イ欄病名の重みは 1、ア欄病名の重みは 0.85 となる。

1556 列目から 1577 列目までは、各付帯情報の有無であり、付帯情報の詳細は表 3 に示すとおりである。また今回の予測タスクの目的変数である仮原死因の変更有無は 1578 列目に記載されている。本研究は、Iris が算出した仮原死因はオートコーディングシステムが算出する仮原死因と同様であるという前提で行っている。

2.2 機械学習手法の詳細

2.3 節で生成したデータを用いて機械学習を行う。機械学習では、表 2 に示す 2〜1577 列目を用いて 1578 列目の変更有無を予測する。

機械学習の手法として、勾配ブースティング決定木(Gradient Boosting Decision Tree: GBDT)の一種である XGBoost を使用する。XGBoost は、アンサンブル学習(複数の弱学習器を組み合わせる手法)のブースティング(弱学習器を逐次的に構築)と決定木(木構造を用いて分類や回帰を行う手法)を組み合わせたものであり、多くの領域で比較的精度の高い結果を算出することができると言われている [7]。ただし、テストデータの大きさやパラメータの設定が精度に影響を与える可能性が高く、パラメータの設定に関して調査が必要である。そこで、本研究では以下のような手順で進めた。

表 1 仮原死因変更・付帯情報の有無別の件数

	付帯情報あり	付帯情報なし
仮原死因変更あり	**11,857** (/320,008≒3.7%)	**16,957** (/320,008≒5.3%)
仮原死因変更なし	**69,838** (/320,008≒21.8%)	**221,356** (/320,008≒69.2%)

表 2 機械学習用データの仕様

列番号	列名	内容
1	age	年齢(小数点第 1 位まで)
2	sex	性別(男:1、女 2)
3 〜 1555	各 ICD-10	存在しない:0 存在する 仮原死因に該当:1 ⅠⅡ欄記載有 : 0.85,0.7,0.55,0.4,0.25
1556 〜 1577	各付帯情報	付帯有りに該当する記載がある:1 ない:0
1578	変更有無	仮原死因が確定原死因と一致する :1 しない:0

表 3 付帯情報の内容

1556	手術の有無
1557	手術の詳細記述
1558	手術に関して備考がある場合はフラグ
1559	手術年月日
1560	解剖の有無
1561	解剖の詳細記述
1562	解剖に関して備考がある場合はフラグ
1563	自然死・不慮の外因死・その他の不詳の死などの種類を表す数字
1564	障害発生年月日
1565	障害発生時分
1566	住居・工場・道路などの障害が発生した場所の種類を表す数字
1567	傷害が発生したところの種別にない場合の記述(その他)
1568	傷害発生都道府県
1569	傷害発生市郡
1570	障害発生区町村
1571	障害発生時の手段及び状況に関する詳細記述
1572	傷害発生に関して備考がある場合はフラグ
1573	妊娠・分娩時における母体の病態または以上に関する詳細記述
1574	「生後 1 年未満での病死」に関して備考がある場合はフラグ
1575	その他に付言すべきことに関する詳細記述
1576	備考欄に外字がある場合はフラグ
1577	備考に関する詳細記述

I. 木の深さを XGBoost のデフォルト値で固定し、テストデータの大きさを以下のように変更
データ全体の（ア） 10％,（イ） 15%,（ウ） 20%,（エ）25%, （オ）30%
テストデータ以外のデータは訓練データとして用いた。

II. 最適な木の深さをグリッドサーチを用いて決定
グリッドサーチとは、学習モデルに用いられるハイパーパラメータ（調整の必要があるパラメータ）を調整する手法であり、指定したハイパーパラメータの全組合せに対して学習を行い、最適なパラメータを選択する。ただし、組み合わせ数が多いほど時間を要するため、今回は決定木の深さにのみ着目し、深さを 4, 6, 8 と変え、最適な深さを決定した。

III. 上記で得られた最適な値を用い学習を実行、交差検証の実行
交差検証では、データをいくつかのセットに分割したのち、1 つをテストデータのセット、残りを学習データとして精度の評価を行う。そして、分割したデータのセットすべてが 1 回ずつテストデータになるように学習を行なって出てきた精度の平均を算出する。今回は、ステップ I でテストデータの割合にて大きく精度が影響を受けないことを確認の上、分割数を 5 とした。

2.3 評価

2.2 節から学習された分類モデルに基づき、精度の算出を行った。また、決定木系の手法である XGBoost では、予測に際してのパラメータの重要度を算出できるため、これを用いて考察を行った。

2.4 倫理面への配慮

本研究では、個人情報に関わる調査・実験は行っていない。研究の遂行に当たっては、各種法令や「人を対象とする医学系研究に関する倫理指針」を含めた各種倫理指針等の遵守に努めた。

3. 結果

3.1 テストデータの大きさ別の結果

各テストデータの大きさ別精度の推移を図 1 に示す。最も精度が高いものは、2.2 節の（イ）15%の 90.43%、また最も低いものは（ア）10%の 90.08%であった。若干の変動はあるものの、テストセットに使用するデータの割合には大きく影響を受けていないことが確認された。そこで、以降の実験では一般的に良く用いられる（ウ）20%の設定を用い、5 分割交差検証をすることとした。

3.2 グリッドサーチと分類精度の結果

グリッドサーチをした結果、最適な木の深さは 4 であった。この事前調査で得られた最適なパラメータを用いて学習を行った結果を図 2 に示す。正解率(Accuracy) は 90.30%であった。図2中の上の表はテストデータの分類件数を四分表にまとめた内訳である。一方、図 2 中の下の表は「変更なし」あるいは「変更あり」に着目した際の感度（recall）・適合度（precision）・F 値・特異度（specificity）の結果である。

一方、学習したモデルにおける変数重要度を図 3 に示す。縦軸の数字は、表 2 に示す機械学習用データの仕様の番号と対応しており、最も影響を与えたのは年齢、次に備考有無、I501（左心不全に関する病名）の有無、その他付言すべき事柄の影響、の順となった。

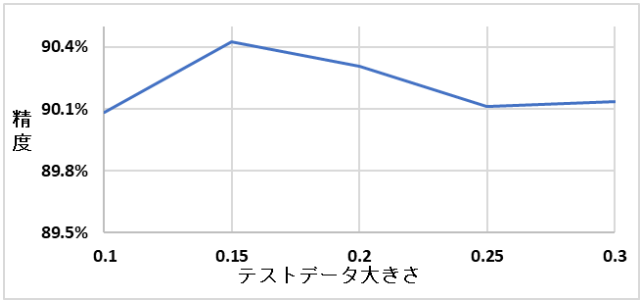
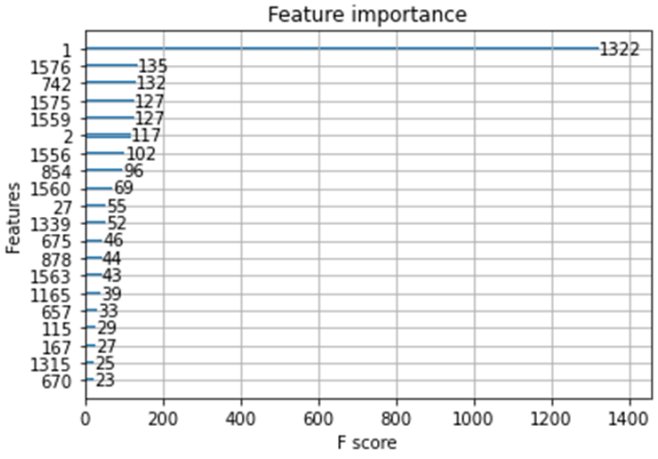


図 1 テストデータの大きさ別の精度推移

		正解		
	変更	なし	あり	計
予測	なし	13545	1161	14706
	あり	423	1209	1632
	計	13968	2370	16338
accuracy		0.90304		

	あり	なし
recall	97.0%	51.0%
precision	92.1%	74.1%
F	0.94	0.60
specitivity	51.0%	97.0%

図 2 5 分割交差検証（木の深さ 4）の結果



1	:	年齢	1339	:	R53（衰弱）
1576	:	備考有無	675	:	I214（急性心内膜下梗塞）
742	:	I501（左心不全）	878	:	J679（過敏性肺炎）
1575	:	その他付言すべき事柄	1563	:	自然死以外
1559	:	手術年月日	1165	:	N185（末期腎不全）
2	:	性別	657	:	I092（慢性リウマチ性心膜炎）
1556	:	手術有無	115	:	C164（胃幽門部癌）
854	:	J188（閉塞性肺炎）	167	:	C348（細気管支肺胞上皮）
1560	:	解剖有無	1315	:	R068（無呼吸発作）
27	:	A415（グラム陰性菌敗血症）	670	:	I208（安静時狭心症）

図 3 変数重要度（上位）
ICD-10 コードの場合は、傷病名例を（）内に表記

4. 考察

本研究での提案手法にて仮原死因の変更有無に関する分類モデルを構築した結果、約 9 割の精度で分類が可能であることが判明した。付帯情報については内容を考慮しない記載の有無のみを用いる最も簡略なベースライン手法であるにも関わらず、高い精度を実現しており、今後付帯情報の内容を加味することでさらなる精度向上が見込めることから非常に有望な手法と考えられた。また、図 1 のテストデータの大きさでの検討で精度に大きな違いがなかったこと、さらにハイパーパラメータの探索で最適とされた結果 (正解率 90.3%, 図 2 参照)も、デフォルト値を用いた正解率(図 1 参照) と大きな違いがなかったことから、 今回の実験ではテストデータの大きさやハイパーパラメータの選択の違いによる学習への影響は少なく、ロバストな結果と考えられる。

また、本研究の過程で得られた、Iris による仮原死因と確定原死因との比較結果（表 1）から、付帯情報がある場合に人手による確認が行われ、実際に原死因コードの変更に至るケースは全体の 14.5%程度である。つまり、仮原死因コードの変更の有無が高精度に自動分類されることにより、人手作業の 8 割を削減し、変更内容の決定に注力することができるようになることから、実業務の効率化に大いに資すると考えられる。

学習されたモデルにおいては図 3 で示されたように年齢・性別や ICD-10 コード、付帯情報有無の項目に関して、変数重要度が大きいことがわかった。また、備考欄の有無やその他付言すべき事柄に記入があるか付帯情報が影響を与えていることもわかった。今回は単純な付帯情報の有無情報のみであったが、自由記載内容の情報も学習させることで、単純な仮原死因の変更有無だけでなく、変更内容の提示にも有効な手段になる可能性がある。

次に、提案手法による予測結果と正解が異なるものについての考察を行う。ただし、本研究では Iris がオートコーディングシステムと同じ挙動であるとの仮定をおいており、仮原死因と確定原死因の違いを手がかりにした機械学習を行っている以上、両者の原死因コード導出仕様の違いが、学習に悪影響を与えている可能性がある。現に表1に示す通り、「付帯情報がない」にも関わらず、Iris が導出した仮原死因と国内の確定原死因が異なっているケースが 5%程度存在している。（これは付帯情報による影響では無いため、両者の原死因決定方法が同一であれば本来存在しないはずである）。これは本研究の限界である。

誤りの中で、機械学習による予測で「変更なし」であったが正解は「変更あり」であった場合(FN: False Negative)は、本提案手法のシステムを実業務に導入する際には、大きな問題となる。なぜならば、自動手法で「確認の必要なし」とされたもののうち見逃しが存在してしまっているからである。このようなケースのうち 108 件（1,161 件中）に関しては、Iris が仮原死因を決定できていなかった。これは、ICD-10コードを付与したのちに Iris で仮原死因付与を行う際に、Ⅰ Ⅱ欄に記載された傷病名に対して付与した ICD-10 コードが日本独自コードである場合や、入力された ICD-10 コードに不備（前期破水は男性には使えないなど、性別や年齢と ICD-10 コードに不一致）がある場合、Iris は仮原死因を付与することができないためである。この問題は、ICD-10 コード付与時に改めて日本独自コードを修正することや Iris が出力するエラーメッセージを参照し仮原死因の付与を行い対策する必要があり、今後の課題である。また、FN の中で各付帯情報の記載がある件数を算出すると、手術欄の詳細記述欄に記載がある場合が最も多い（686/1161 件）ことがわかった。今後より精度を上げるためには、詳細記述欄の記述内容に着目する必要があり（特定の単語や文言が仮原死因変更に影響を与える可能性があるため）、今後の課題である。

一方、機械学習による予測で「変更あり」であったが正解は「変更なし」であった場合 (FP: False Positive) に関しても手術欄の詳細記述欄そして手術の有りの番号にマークがされている場合が最も多かった（321/423 件）。FN の場合と同様に付帯情報の詳細記述欄などを参照する必要がある。ただし、実用する際は必ず人の目視確認が入るため、数は減らすべきであるが、FN のケースを優先して改善していく必要がある。

以上より、本研究で用いた付帯情報の影響による仮原死因の変更有無の判別手法は有望であり、今後 BERT 等を用いた付帯情報の内容を加味した手法を開発する上でベースラインとする予定である。

5. 結論

本研究では、付帯情報の影響による仮原死因の変更有無が判別可能かの調査を行った。その結果、機械学習ベースの分類モデルの XGBoost を用いた本手法による変更予測モデルの精度は、約 9 割であった。重要度の高い因子は、年齢、備考欄の記載の有無、I501コードの存在、その他不言すべき事柄の記載の有無、手術年月日の有無などであった。記載の有無だけを用いたベースライン手法でも、高い精度だったため、本研究で用いた手法は有用であることが示唆された。

今後は、自然言語記載内容の分散表現の利用等による、さらなる精度向上や、原死因を変更する場合の変更後のコード提示に取り組む予定である。

謝辞

本研究は、厚生労働科学研究費補助金(政策科学総合研究事業（統計情報総合研究事業））「死因統計の精度及び効率性の向上に資する機械学習の検討に関する研究（19AB1003）」の一環として実施したものである。
本研究に関し、開示すべき利益相反はない。

参考文献

1) 厚生労働省大臣官房統計情報部編. 疾病、傷害および死因統計分類提要 ICD-10 準拠第 2 巻 Instruction manual（総論）. 厚生労働統計協会, 2016.
2) 今井健, 明神大也, 大井川仁美, 香川璃奈, 今村知明. 原死因確定作業についての実態・問題点の把握, ならびに正確・効率性向上に向けた機械学習の適用可能性と課題に関する調査研究. 厚生の指標 2020 ; 67(3) :17-24.
3) 明神大也, 大井川仁美, 香川璃奈, 今村知明, 今井健. 死因統計の精度と効率性の向上に向けた我が国の原死因確定課題の抽出. 医療情報学連合大会論文集 2020 ; 40 :677-682.
4) Iris Institute. Federal Institute for Drugs and Medical Devices. [https://www.dimdi.de/dynamic/en/classifications/iris-institute/index.html (cited 2021-Aug-30)]
5)大井川仁美, 明神大也, 香川璃奈, 今村知明, 今井健. 原死因確定プロセスにおける IRIS の国内導入可能性に関する基礎的な検討. 医療情報学連合大会論文集 2020 ; 40 :677-682.
6) 標準病名マスター作業班. 病名くん 2.0. 標準病名マスター作業班. [http://www.byomei.org/wg/index.html (cited 2021-Aug-30)].
7) Chen, T, Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 ; 785-794