

売り切れ情報を考慮したマルチタスク学習に基づく 惣菜の需要予測モデルに関する研究

情報数理応用研究

5220C004-0 上原諒介

指導教員 後藤正幸

A Study on a Demand Forecasting Model for Side Dish Based on Multi-task Learning Considering Sold Out Information

UEHARA Ryosuke

1. はじめに

コロナ禍の影響も相まって、スーパーマーケットチェーンの惣菜需要は高まりをみせている。惣菜は消費期限が短く、また見込み生産商品のため廃棄が生じやすい。そのため、惣菜の高精度な需要予測は、在庫管理の観点や、昨今求められているSDGsの観点からも重要な取り組みである。

本研究では、スーパーマーケットチェーンにおける購買履歴データから、惣菜の需要量の予測を行う機械学習モデルを構築し、小売店の需要予測における有用な知見を得ることを目的とする。一般に、顧客の購買行動における需要変動要因として、季節性や曜日効果などの周期的な要因に加え、天候による嗜好の変化や来客数の増減などの短期的な要因が挙げられる。さらに、惣菜商品では顧客の目的の惣菜が次品していた場合に、別の惣菜を代わりに購入する行動（このように生じる商品需要を本研究では代替需要と呼ぶ）を無視できない。そしてこの代替需要は、惣菜の品揃えや割引率などにも影響を受ける。そこで本研究では、これらの需要変動要因を考慮した上で、多種のアイテムの需要の連動性を反映しながら全アイテムの需要を定量的に予測するモデルを構築する。

一方、本研究で対象とするスーパーマーケットチェーンでは、各店舗でマネージャーの経験則に基づいた定性的な需要予測を行い、各アイテムの出品数を決定している。しかし、出品数に対して実需用の方が大きかった場合、売り切れが生じてしまう。このとき購買履歴データにおいては、実際に販売された数量しか観測することができないため、アイテムの実需要が観測されていないこととなる。このように、従属変数の一部の値が制限されているデータを打ち切りデータと呼ぶ。打ち切りデータにおいて、値域の制限のない通常のデータと同様にパラメータ推定を行った場合、偏りのある推定結果が得られてしまう。これに対し、打ち切りデータの分布の偏りを考慮した上で、回帰モデルのパラメータ推定が可能なモデルとして Tobit モデル [1] がある。本研究では Tobit モデルを導入することにより、アイテムの売り切れを考慮した上でパラメータ推定を行う。

さらに、スーパーマーケットの特性上、出品管理のためには単一アイテムだけの予測を行うのではなく、陳列されている全惣菜アイテムの個々の需要を把握する必要がある。そこで、マルチタスク学習 [2] の概念を援用し、多種のアイテムの需要量を同時に目的変数としたモデルを構築することを考える。Tobit モデルを適用し、マルチタスク学習を実行した研究に Sun らの研究 [3] がある。Sun らは、複数の目的変数を予測する際、パラメータの更新を行う 1 つの構成単位（以下、Block と呼ぶ）を複数積み重ね、各 Block でパラメータを最適化するネット

ワークモデルを提案した。そして、このモデルを医療用の脳画像データに対して適用し、0 以下の値が観測されない目的変数の予測を行っている。このとき、各 Block の出力の活性化関数に Rectified Linear Unit (以下 ReLU 関数) [4] を用いることで、下側打ち切りのデータに対して高精度な予測を実現した。しかし本研究では、目的変数となるアイテムの需要量は出品数を上限とした上側打ち切りのデータであることに加え、日にちやアイテムごとに需要量が大きく変動し、打ち切りの上限値も異なるため、Sun らのモデルを適用することは適切ではない。

そこで本研究では、これらの問題に対応しつつ、Block を積み重ねて予測精度の向上を図ったモデルにより、複数の惣菜アイテムの需要を一括予測する手法を提案する。具体的には、単に Block を積み重ねるのではなく、勾配ブースティング [5] を援用し、前の Block で学習が上手くいかなかった部分を適切に学習できるように次の Block を学習させる手法を新たに構築する。これにより、需要が大きく変動し予測が困難なアイテムに対しても適切に学習を行うことが可能となる。そして各 Block の学習には、入力層、中間層、出力層で構成される 3 層の単純な Neural Network (以下、NN) を用いて、多種のアイテムの関係を捉えた予測を行う。このようなモデルをスーパーマーケットチェーンの店舗ごとに構築し、店舗の特性を考慮した需要予測モデルを構築する。最後に、提案手法をスーパーマーケットチェーンの実際の購買履歴データに適用し、明らかとなる結果について考察を与える。

2. 準備

2.1. Tobit モデル [1]

Tobit モデル [1] は、Tobin によって提案された、打ち切りデータのパラメータを推定する回帰モデルである。全データを N 、目的変数を y_i 、説明変数ベクトルを x_i 、打ち切りがなかった場合の真の目的変数を y_i^* とする ($i = 1, \dots, N$)。目的変数の 0 以下の値が観測されない状況であった場合、以下のような関係性となる。

$$y_i^* = x_i^T \gamma + \epsilon, \quad (1)$$

$$y_i = \begin{cases} y_i & y_i^* > 0, \\ 0 & y_i^* \leq 0. \end{cases} \quad (2)$$

ただし、 γ は回帰モデルの係数ベクトル、 ϵ は正規分布に従う誤差項であり、 $\epsilon \sim \mathcal{N}(0, \sigma^2)$ である。式 (2) において、Tobit モデルでは、 $x_i^T \gamma + \epsilon$ に等しい未観測の潜在変数の存在を仮定している。このとき、確率モデルは以下のように書き表すことができる。

$$\Pr(y_i, \mathbf{x}_i | \gamma, \sigma^2) = \left\{ \frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\gamma}}{\sigma} \right) \delta(y_i > 0) + \Phi \left(\frac{0 - \mathbf{x}_i^T \boldsymbol{\gamma}}{\sigma} \right) \delta(y_i \leq 0) \right\}. \quad (3)$$

ただし、 $\phi(\cdot)$ は $\mathcal{N}(0, 1^2)$ の標準正規分布に従う確率密度関数を、 $\Phi(\cdot)$ はその累積分布関数を表し、 $\delta(\cdot)$ は条件を満たすとき 1、さもなければ 0 をとる指示関数である。

全 N 件のデータに対する対数尤度関数 $\mathcal{LL}(\boldsymbol{\gamma})$ は以下で表される。

$$\mathcal{LL}(\boldsymbol{\gamma}) = \sum_{i=1}^N \log p_{\boldsymbol{\gamma}}(y_i, \mathbf{x}_i). \quad (4)$$

Tobit モデルにおいては、以上の対数尤度関数を最大化するようなパラメータ $\boldsymbol{\gamma}$ を最尤推定する。

2.2. Sun らのモデル [3]

Sun ら [3] は、ネットワークモデルを用いて、複数の目的変数を予測するモデルを提案した。パラメータ更新を行う構成単位である Block を複数積み重ねることにより精度の向上を図ったモデルで、各 Block の損失関数には Tobit モデルと同様の損失関数を用いている。Sun らのモデルのイメージを図 1 に示す。

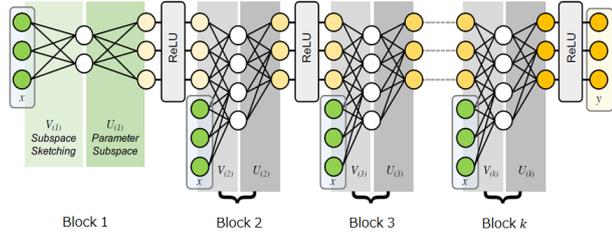


図 1: Sun らのモデルのイメージ図 [3]

1 つ目の Block では、事前学習された重み行列の行列分解を行い、2 つ目以降の Block では前の Block の出力を元の入力と結合したものを新たな入力として、勾配降下法でパラメータを求めている。

D 次元の説明変数ベクトルを $\mathbf{x}_i \in \mathbb{R}^D$ 、 T 次元の目的変数ベクトルを $\mathbf{y}_i \in \mathbb{N}^T$ とすると、 N 件の観測データの組み合わせは $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ と表せる。

各予測値に対し、非線形関数の ReLU 関数を適用する。ReLU 関数は入力値が 0 未満のときに 0 を出力し、入力値が 0 以上のときに入力値をそのまま出力する関数である。Sun らの研究で対象とするデータでは、目的変数は 0 未満の値が観測されないため、ReLU 関数によってうまく打ち切りを表現することができる。

$$y_{i,t} = \text{ReLU}(\mathbf{W}_t^T \mathbf{x}_i + \epsilon). \quad (5)$$

ここで、 $\mathbf{W} = [\mathbf{W}_1, \dots, \mathbf{W}_T]^T \in \mathbb{R}^{T \times D}$ は事前学習された重み行列、 ϵ は正規分布に従う誤差項であり、 $\epsilon \sim \mathcal{N}(0, \sigma^2)$ である。このとき \mathbf{W} に対し、低ランクの部分空間 \mathbf{U} と埋め込み係数 \mathbf{V} を用いて、 $\mathbf{W} = \mathbf{U}\mathbf{V}$ という関係性を仮定する。ただし、 $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_T]^T \in \mathbb{R}^{T \times R}$ 、 $\mathbf{V} \in \mathbb{R}^{R \times D}$ である。

このとき、確率モデルは以下のように表される。

$$\Pr(\mathbf{y}_i, \mathbf{x}_i | \mathbf{U}, \mathbf{V}, \sigma^2) = \prod_{t=1}^T \left\{ \phi \left(\frac{y_{i,t} - \mathbf{U}_t^T \mathbf{V} \mathbf{x}_i}{\sigma} \right) \delta(y_{i,t} \in (0, \infty)) + \Phi \left(\frac{0 - \mathbf{U}_t^T \mathbf{V} \mathbf{x}_i}{\sigma} \right) \delta(y_{i,t} = 0) \right\}. \quad (6)$$

ただし、 $\phi(\cdot)$ は $\mathcal{N}(0, 1^2)$ の標準正規分布に従う確率密度関数を、 $\Phi(\cdot)$ はその累積分布関数を表し、 $\delta(\cdot)$ は条件を満たすとき 1、さもなければ 0 をとる指示関数である。

Sun らは、複数の Block を積み重ね、各 Block で目的関数を学習させることにより、より高い精度を得られることを示している。

k 番目の Block におけるネットワーク、部分空間、埋め込み係数をそれぞれ $f_{[k]}(\cdot)$ 、 $\mathbf{U}_{[k]}$ 、 $\mathbf{V}_{[k]}$ と表すと、 k 番目の Block における出力は以下のように表すことができる。

$$f_{[k]}(\mathbf{x}_i) = \text{ReLU}(\mathbf{U}_{[k]} \mathbf{V}_{[k]} [f_{[k-1]}(\mathbf{x}_i); \mathbf{x}_i] + \epsilon). \quad (7)$$

2.3. 勾配ブースティング [5]

勾配ブースティング [5] はアンサンブル学習の一つで、複数の学習器を直列に学習させ、直前までの学習器の予測誤差を小さくするよう次の学習器で学習を行うモデルである。 \mathbf{x} をもとに \hat{y} を出力するモデルを $F^*(\mathbf{x})$ と表した時、 $F^*(\mathbf{x})$ は損失関数 $L(y, F(\mathbf{x}))$ を最小化するように学習される。

$$F^*(\mathbf{x}) = \arg \min_{F(\mathbf{x})} E_{y, \mathbf{x}} L(y, F(\mathbf{x})). \quad (8)$$

勾配ブースティングでは、 $F^*(\mathbf{x})$ を最小化するために、逐次的にモデル ($k = 1, \dots, K$) を加算しながら、学習を行う。

$$F(\mathbf{x}) = \sum_{k=1}^K \beta_k h(\mathbf{x}; \mathbf{a}_k). \quad (9)$$

ただし $\beta_k (k = 1, \dots, K)$ は展開係数であり、関数 $h(\mathbf{x}; \mathbf{a}_k)$ はパラメータ \mathbf{a} を持つ \mathbf{x} の関数を表す。

3. 提案モデル

3.1. 着想

本研究で対象とするスーパーマーケットチェーンの購買履歴データに対し、Sun らのモデルを適用し、売り切れを考慮したうえで多種アイテムの需要を一括予測するモデルを考える。しかし本研究で対象とするデータは、目的変数となるアイテムの需要量は出品数を上限とした上側打ち切りのデータであるため、ReLU 関数の当てはまりが特別良いわけではない。さらに、日にちやアイテムごとに需要量が大きく変動し、打ち切りの上限値も異なるため、Sun らのモデルを適用することは適切ではない。そこで本研究では、これらの問題に対応しつつ、Block を積み重ねて予測精度の向上を図る。具体的には、単に Block を積み重ねるのではなく、前の Block での学習が上手くいかなかった部分を適切に学習できるように次の Block を学習させる手法を提案する。

3.2. 概要

スーパーマーケットチェーンの購買履歴データを用いて、多種のアイテムの需要量を目的変数とした予測モデルを構築する。

本研究では、複数のアイテムを同時に目的変数とし、入力層、中間層、出力層で構成される3層の単純なNNを1つのBlockとして複数Blockを積み重ねる。このとき、売り切れが生じたために実需要が観測されていない場合を考慮し、各Blockでは、Tobitモデルの損失関数を最小化するように学習を行う。このとき、勾配ブースティングの概念を援用し、前のBlockの出力を元に次のBlockのパラメータの学習を行うモデルを提案する。このモデルにより、前のBlockで適切に学習されなかった箇所を次のBlockで重点的に学習させるようBlockを積み重ねることができる。ここで、アイテムの需要量は、各店舗の特性に大きく影響を受けると考えられることから、このようなモデルを店舗ごとに構築する。さらに廃棄履歴データをもとに、閉店時の残在庫数が0となった場合、すなわち廃棄が生じていない場合に売り切れが生じていると捉える。提案モデルのイメージを図2に示す。

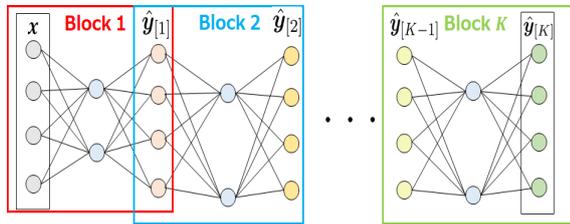


図2: 提案モデルのイメージ

3.3. 学習アルゴリズム

2.1節、2.2節に則り、変数を定義する。 D 次元の説明変数ベクトルを $\mathbf{x}_i \in \mathbb{R}^D$ 、需要量を表す R 次元の目的変数ベクトルを $\mathbf{y}_i \in \mathbb{R}^R$ とすると、 N 件の観測データの組み合わせは $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ と表せる。ここで、店舗 s におけるアイテム r の第 i 日目の需要量 $y_{s,i,r}$ は、閉店時に売り切れが生じていないときは真の需要量がそのまま観測されるが、売り切れが生じている場合には、データに打ち切りが発生している。真の需要量を $y_{s,i,r}^*$ 、閉店時の残在庫数を $c_{s,i,r}$ とすると、需要量を以下のように表すことができる。

$$y_{s,i,r}^* = \begin{cases} y_{s,i,r} & c_{s,i,r} > 0, \\ y_{s,i,r} + \alpha & c_{s,i,r} = 0. \end{cases} \quad (10)$$

ただし、 α は非観測な需要量を表す ($\alpha > 0$)。

各Blockでの出力 $\hat{\mathbf{y}}_{[k]}$ は以下のように表される。

$$\hat{\mathbf{y}}_{[k]} = \begin{cases} f_{[1]}(\mathbf{x}, \boldsymbol{\theta}_{[1]}) + \epsilon, & k = 1, \\ f_{[k]}(\hat{\mathbf{y}}_{[k-1]}, \boldsymbol{\theta}_{[k]}) + \epsilon, & k \geq 2. \end{cases} \quad (11)$$

ただし、関数 $f_{[k]}(\cdot)$ は k 番目のBlockにおけるネットワーク、 $\boldsymbol{\theta}_{[k]}$ は k 番目のBlockにおけるNNのパラメータを表し、 $\epsilon \sim \mathcal{N}(0, \sigma^2)$ である。

このとき、確率モデルは以下のように書き表すことができる。

$$\Pr(\mathbf{y}_{[k]s,i} | \mathbf{x}_{s,i}, \sigma^2) = \prod_{r=1}^R \left\{ \frac{1}{\sigma} \phi \left(\frac{y_{s,i,r} - \hat{y}_{[k]s,i,r}}{\sigma} \right) \delta(c_{s,i,r} > 0) + \Phi \left(\frac{y_{s,i,r} - \hat{y}_{[k]s,i,r}}{\sigma} \right) \delta(c_{s,i,r} = 0) \right\}. \quad (12)$$

ただし、 $\phi(\cdot)$ は $\mathcal{N}(0, 1^2)$ の標準正規分布に従う確率密度関数を、 $\Phi(\cdot)$ はその累積分布関数を表し、 $\delta(\cdot)$ は条件を満たすとき1、さもなければ0をとる指示関数である。全 N 件のデータに対する対数尤度関数 $\mathcal{LL}(\boldsymbol{\theta}_{[k]})$ は以下で計算できる。

$$\mathcal{LL}(\boldsymbol{\theta}_{[k]}) = \sum_{i=1}^N \log p_{\boldsymbol{\theta}_{[k]}}(\mathbf{y}_{[k]s,i} | \mathbf{x}_{s,i}). \quad (13)$$

本研究では、各Blockで対数尤度関数 $\mathcal{LL}(\boldsymbol{\theta}_{[k]})$ を最大化するようなパラメータ $\boldsymbol{\theta}_{[k]}$ を最尤推定する。

4. 実データを用いた分析

提案手法の有効性を示すため、提案手法を実データに適用し、分析を行う。

4.1. 対象データ

本研究では、東海地方を中心にスーパーマーケットチェーンを展開する株式会社バローから提供された、スーパーマーケットの販売、生産、廃棄履歴データを対象とする。このスーパーマーケットチェーン全体では、1,000種類以上の惣菜アイテムが取り扱われているものの、各店舗で販売されるアイテムは一日数十種類程度であり、日々提供される惣菜の種類も異なる。

また、説明変数として導入する気象データには、日本気象協会から提供された位置情報付き1時間単位の気象データを用いる。使用する変数は気温、日射量、風速、風向、降水量、降雪量、相対湿度、天気であり、位置情報をもとに各店舗でスケールして使用する。

4.2. 分析条件

対象期間は2018年1月2日～2019年12月31日、対象店舗数は207店舗とした。2年間継続的に売られているアイテムを対象とするため、各店舗で年間30日以上売られていた50アイテムを対象惣菜品とし、2018年を学習データ、2019年をテストデータとした。推定精度の評価指標としてMAE(Mean Absolute Error)を用い、アイテムや店舗の全てに関して平均化したMAEを本研究ではMAEと呼ぶ。

4.3. 分析結果

4.3.1. Block数に関する分析

本節では、提案モデルの予測精度を検証する。予測時点での気象データ、全アイテムの在庫数、目的アイテムの予測時点までの販売数をもとに、その以降の需要量を予測する。15時および17時時点を予測時点とした時のBlock数ごとのMAE(テストデータ)を表1に示す。

表1: Block数ごとのMAE

Block数	1	2	3	4	5
15時	29.4	29.4	29.4	29.6	29.5
17時	25.3	23.2	23.3	23.1	22.3
Block数	6	7	8	9	10
15時	29.5	29.0	29.2	29.5	29.1
17時	22.9	22.6	22.8	22.3	22.6

最も低い値を示した箇所を太字で示している。表1より、15時時点の予測ではBlock数が7の時に最も低い値を示し、17時ではBlock数が5の時に最も低い値を示した。

ここで、複雑な実データにおいて、一部の店舗はBlock数を積み重ねることによって予測精度が大きく劣化する可能性が

ある。17時時点の予測において、Block 数が1の時と比較し、提案手法によって MAE が悪化した店舗は 71 店舗あり、全店舗の約 34%にあたる。逆に残りの約 66%の店舗では、Block 数が増加することによって MAE は改善している。そのため、表 1 の結果は、Block 数の増加に伴って推定精度が劣化する一部の店舗の MAE の増加に起因することが分かる。

そこで、店舗ごとに MAE が最小となる Block 数を選択することを考える。店舗ごとに適切な Block 数を選択した際の MAE を表 2 に示す。比較手法は Block 数が 1 であるモデル、すなわち単純な 3 層の NN モデルとした。

表 2: 店舗ごとに適切な Block 数を選択したときの MAE

予測時点	12	13	14	15
比較手法	33.8	30.5	29.8	29.4
提案手法	30.6	27.6	25.8	24.5
予測時点	16	17	18	
比較手法	27.2	25.3	21.9	
提案手法	22.4	19.4	15.3	

表 2 より、すべての時間帯において、提案手法の方が良い精度を示していることが分かる。また、遅い時間帯になるにつれて精度が高まっていることが確認できる。

4.3.2. 店舗・アイテムに関する分析

本節では、Block 数を積み重ねることによって大きく精度が向上した千曲店の分析を行う。千曲店におけるえび天ぷら、海老かき揚げの Block 数ごとの MAE を以下に示す。

表 3: 千曲店におけるアイテムごとの MAE(12 時)

Block 数	1	2	3	4	5
えび天ぷら	21.5	32.5	22.8	32.5	12.0
	6	7	8	9	10
	30.4	11.0	15.8	17.7	32.5
Block 数	1	2	3	4	5
海老かき揚げ	12.8	27.3	14.5	17.1	7.35
	6	7	8	9	10
	5.86	23.4	7.93	27.3	27.3

表 3 より、アイテムによっても適切な Block 数は異なり、Block 数によって MAE が高低下することが分かる。これは、予測精度の高いアイテム群が Block ごとに異なる可能性を示唆している。そこで、アイテムの Block 数ごとの予測精度から、商品の特性を分析することを考える。例として、最も良い精度を示した Block 数が 3, 7 であったアイテムを以下に示す。

表 4: Block 数が 3, 7 の時に最も良い精度を示したアイテム

Block 数 : 3	Block 数 : 7
いも天、うずら串フライ、3種のサーモン丼、まるごとイカ唐揚げ、かつとじ	えび天ぷら、ひとくちヒレカツ、寿司サンド、白めし、ふんわり屋台風、お好み焼き (豚)

表 4 より、Block 数が 3 の時に精度が良いアイテムは揚げ物が多く、Block 数が 7 の時に精度が良いアイテムはご飯ものが多いことが分かる。これらのアイテムは、需要変動要因が類似している可能性が高く、需要変動性の高いアイテム同士であると言える。そのため提案モデルは、アイテムの需要予測に関する特性を捉えるための手段となりえる。

5. 考察

本研究では、3 層の単純な NN を 1 つの Block として、前の Block の出力を踏まえたうえで次の Block の学習を行うモデルを店舗ごとに構築した。一部の店舗では単純な NN を用いた予測が適している一方、多くの店舗では、Block 数を積み重ねることにより予測精度が向上することが示された。このような特徴を考慮し、提案モデルを用いて店舗ごとに異なる Block 数を選択することで、店舗特性を考慮した高い精度での需要予測が可能となる。

また、提案モデルにより需要が相互に依存し合うアイテムのグループを Block ごとに抽出できることを示した。提案モデルでは、抽出されたアイテムのグループをもとに、需要変動性を反映したうえでアイテムの需要予測を行っているため、高い予測精度を示したと考えられる。提案モデルを用いて各店舗のアイテムごとに適切な Block 数を選択することで、アイテムの需要変動性に応じたクラスタリングを行うことができる。

6. まとめと今後の課題

本研究では、スーパーマーケットチェーンの購買履歴データにおいて、売り切れが生じていることに着目し、Tobit モデルを適用した。そして Tobit モデルを学習する NN を 1 つの Block として複数 Block 積み重ね、複数の目的変数を予測するモデルを提案した。提案モデルでは、前の Block の学習がうまくいかなかった箇所を次の Block で重点的に学習するように積み重ねることで、予測精度の向上を図った。

実データ分析では、提案手法を小売チェーンの実データに適用し、予測精度の向上を確認することができた。そして、店舗・アイテムごとに Block 数を選択することで、需要変動性の高いアイテムのグループを抽出できることを示した。今後の課題として、Block 数の決定方法の検討や値引き情報の導入が挙げられる。

謝辞

貴重なデータを提供していただいた株式会社パローおよび日本気象協会に心より感謝申し上げます。

参考文献

- [1] James Tobin. Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, pp. 24–36, 1958.
- [2] Rich Caruana. Multitask learning. *Machine learning*, Vol. 28, No. 1, pp. 41–75, 1997.
- [3] Mengying Sun, Inci M Baytas, Liang Zhan, Zhangyang Wang, and Jiayu Zhou. Subspace network: deep multi-task censored regression for modeling neurodegenerative diseases. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2259–2268, 2018.
- [4] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 315–323. JMLR Workshop and Conference Proceedings, 2011.
- [5] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.