

## 長時間バスロケーションデータに対するサポートベクトル回帰及び 畳み込みニューラルネットワークによる到着時刻予測

### Bus Arrival Prediction using Support Vector Regression and Convolutional Neural Network methods for Long-term Bus Location Data

山崎将貴<sup>1</sup> 名倉 陽太<sup>1</sup> 荒木大輝<sup>2</sup> 豊木博泰<sup>3</sup>  
Shoki Yamazaki Yota Nagura Daiki Araki Hiroyasu Toyoki

#### 1. はじめに

バス運行に関する Web 情報サービスにバスの現在位置情報(以下、「バスロケ」)を付加したサイトが増加している。山梨県にバス県域全体の路線におけるバス位置情報を収集し、バス利用者に位置情報をリアルタイムで提供する Web サービス「やまなしバスコンシェルジュ[1]」がある。周知のとおり、道路状況に影響されるバスは定時運行が困難であり、到着予測時刻のサービスには強いニーズがある。すでにバスロケを用いた到着時刻予測に関する研究は多数行われており[2,3]、実用的なサービスサイトも増えている。しかし、実用サービスにおける精度は必ずしも満足できるものではなく、最近でも精度向上のための研究が続いている。手法としては、ARIMA(Autoregressive Integrated Moving Average)などの時系列分析手法を用いたもの[4]、ニューラルネットワークを用いたもの[5]、k-最近傍法やSVR(サポートベクトル回帰)を用いたものなどがある[6]。これまでの多くの研究は、運行頻度が高い大都市圏のバス運行データに関するものであり、実用サービスも大手バス事業者によるものがほとんどである。しかし、到着(あるいは遅延)予測は、むしろ運行回数が日に数本程度の地方路線バスに対するニーズの方が高いと考えられる。乗り遅れた場合の不利益が大きいため、地方の路線ほど予測が難しい。地方では通勤に自家用車を利用する割合が高く、通勤時間帯と日中との渋滞状況の差が激しい。後述するように、渋滞は日々同様に発生するのではなく、日による揺らぎが大きいため、バスの遅延の日格差が大きいのである。

本研究では、長期にわたって集積したバスコンシェルジュのバスロケデータを用いた線形回帰モデル(LR)、サポートベクトル回帰、k-最近傍法、リッジ回帰モデル、畳み込みニューラルネットワークによる予測精度の比較を行った。又、予測時の道路状況を反映させるために VICS データを CNN 以外の手法に利用し、予測精度の比較を行った。VICSは Vehicle Information and Communication System の略称であり、渋滞や交通規制などの道路交通情報を、リアルタイムにカーナビに届けるシステムである。

#### 2. 線形単回帰による予測とボトルネック

はじめに、予測対象地点到着時刻を目的変数とした単回帰による分析を行う。各停留所で予測対象地点に対して回帰直線を算出し、予測をしている。なお、平日、休日ともに運行している場合は平日と休日で分類分けを行いモデルに適用した。2路線に関する単回帰モデルによる予測結果を図1(路線1)と図2(路線2)に示す。縦軸が $\pm 2$ 分未満で予測できた割合、的中率(Hitting Percentage)である。横軸は始発を1番目としたバス停の通し番号であり、予測値を算出した地点(バス停)である。両路線とも予測対象地点は終点

とした。路線1(図1)は鵜沢 07:07 発~甲府駅 08:03 着の路線である。甲府盆地西南部から釜無川を渡って盆地中北部の甲府駅に至る山梨交通の路線であり、橋の通過が通勤時間帯に重なりそこが典型的なボトルネックになっている。路線2(図2)はほぼ同じ時間帯に盆地南部から北上し甲府駅に至る豊富 07:15 発~県立中央病院 08:16 着の路線である。両路線とも、朝の通勤時間帯の便であり、予測対象地点(終点)では10分以上の遅れが5割以上ある非常に遅れる路線である。路線1の平日の的中率と路線2の的中率を比較すると両者には大きな差が見られる。路線1では的中率が30%を超えてくるのは20番目の停留所を超えたあたりからであり、路線2では始発の地点からでも40%程度の的中率である。遅延状況は両者とも同じであるのに対し、路線1は予測しにくく、路線2は予測しやすい結果となった。この原因は路線1のボトルネック部分にある。ボトルネック部分では渋滞が日によって不規則に発生するため、同じ曜日でもバスの運行速度にばらつきが生じる。このばらつきが予測を困難にし、精度を下けている。路線1のようなボトルネックを抱える路線では、各バス停通過時刻のデータだけで予測をすることは難しい。精度向上のため、本研究ではいくつかの機械学習の適用と VICS データの利用について検証した。

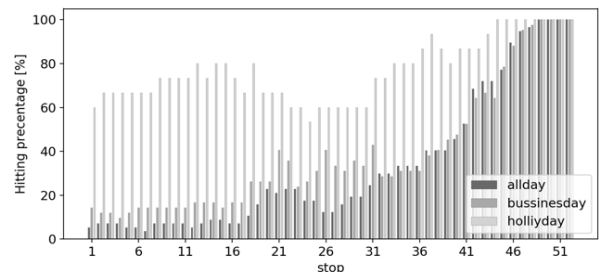


図1 鵜沢 07:07 発~甲府駅 08:03 着線の的中率(単回帰)

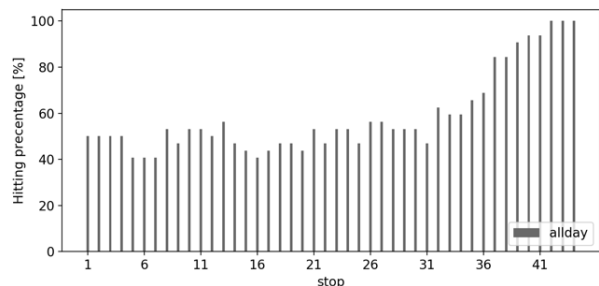


図2 豊富 07:15 発~県立中央病院 08:16 着線の的中率(単回帰)

- 1 山梨大学大学院医工農学総合教育部電気電子コース
- 2 山梨大学工学部電気電子工学科
- 3 山梨大学大学院総合研究部工学域

### 3. VICS データの利用方法と前処理

VICS はルート検索や渋滞回避などに用いられており、(財)道路交通情報通信システムセンターが収集、処理、送信などを行っている。また、公益財団法人日本道路交通センター(JRTIC)より、過去の VICS データを加工した CSV 形式のデータが提供されている。本研究では山梨県内の過去の VICS データを 2017 年 4 月から 2017 年 10 月までの 7 ヶ月分を利用した。構成は表 1 となっており、5 分刻みのデータとなっている。

VICS センサは交通量の多い地点に設置されているが不均一であるため、予測する路線に適した道路情報を取得する必要がある。バスロケデータと VICS データの関連づけを以下に述べる。

バスコンシェルジュの路線情報は緯度経度の列(折れ線状)として格納されている(図 3 の赤線部分)。以下では各緯度経度の接点を経路セグメントと呼ぶ。一方、VICS データは上流部分と下流部分の 2 つのセンサで 1 セットのデータとなっており、それぞれの位置情報が格納されている。図 3 に示すように経路セグメントを中心とした半径 100m 以内の距離にある VICS センサをその経路セグメントに所属させ、所属した VICS センサと対応する VICS センサが他の経路セグメントで検出された場合、その区間を利用可能な区間とした(図 3 における S1-S2 区間)。また、方向や誤検出の確認を行い、不具合があった箇所は修正して利用した。

予測に使用するデータは VICS データと各バス停通過時刻データ、そして平日と休日(祝日を含める)のダミー変数である。VICS データは予測する地点(バス停通過地点)から予測対象地点までの区間で利用できるデータを抽出し、利用する。図 4 に示すケースでの予測のモデル式は以下のようになる。

$$P_E = f\{t_A, D_b, D_h, V_n(t_A), V_n(t_A - \tau), V_n(t_A - 2\tau), V_{n+1}(t_A), V_{n+1}(t_A - \tau), V_{n+1}(t_A - 2\tau)\}$$

$P_E$ : バス停 E の予測到着時刻  
 $t_A$ : バス停 A の到着時刻  
 $D_b, D_h$ : 平日, 休日のダミー変数

ここで、 $V_n(t_A)$  は VICS データから得られる値であり、区間 n の時刻  $[t_A - \tau, t_A]$  での渋滞長  $U$  [m] を表す。ただし、 $\tau = 5$  [min] である。

$$V_n(t_A) = \begin{cases} U \\ 0 \text{ (データなし)} \end{cases}$$

$U$ : 渋滞長(m)

VICS データは 5 分毎のデータとなっており、渋滞が発生していなければ、その間のデータは存在しない。そのため、バス停通過時刻  $t_A$  から、5 分毎に過去 15 分内の VICS データを参照し、データが存在している場合は VICS データ内の渋滞長  $U$ 、存在しなければ 0 を  $V_n(t_A)$  の値とした。

また、VICS データの入力を各区間で 1 つにした、直近データのみ、単純平均、荷重平均の 3 パターンに加え VICS データを入力に利用しない場合の検証も行った。

データの前処理の方法を図 5 に示す。検証データはバス停通過時刻と VICS データに加えて予測地点の到着時刻の

実測値が格納されている。バス停通過時刻と予測地点の到着時刻の実測値は、計算を簡易化するため、時刻を秒数に変換した値を用いた。まず、検証データをトレーニングデータとテストデータに分割する。割合は 7 : 3 とした。その後、検証データに対して正規化を行う。本研究では正規化は代表的な min-max 正規化法を用いる。

表 1 VICS データの詳細

列項	項目説明
A	提供日時 ※年/月/日 時:分
B	都道府県コード番号 ※JIS X 0402 に基づく
C	経路番号 ※JARTIC が独自に付した 8 桁の番号(非公開)
D	路線名称
E	方向
F	下流端点の緯度 ※日本測地系 2000、度分秒の 6 桁で表示
G	下流端点の経度 ※日本測地系 2000、度分秒の 7 桁で表示
H	下流高さ ※-1:入力なし
I	下流地点名称
J	下流地点整列コード ※JARTIC が独自に付した 3 桁の番号(非公開)
K	上流端点の緯度 ※日本測地系 2000、度分秒の 6 桁で表示
L	上流端点の経度 ※日本測地系 2000、度分秒の 7 桁で表示
M	上流高さ ※-1:入力なし
N	上流地点名称
O	上流地点整列コード ※JARTIC が独自に付した 3 桁の番号(非公開)
P	地点間距離(m)
Q	原因コード ※-1:入力なし
R	規制コード ※-1:入力なし
S	渋滞及び混雑区分 ※1:渋滞、2:混雑、-1:入力なし
T	下流端点までの距離(m) ※0:下流端点の位置、整数::下流端点の進行方向 手前、負の数:下流端点の進行方向先
U	渋滞及び混雑長(m)

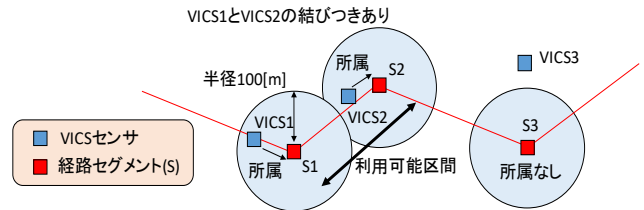


図 3 VICS センサと路線情報の紐付け方法

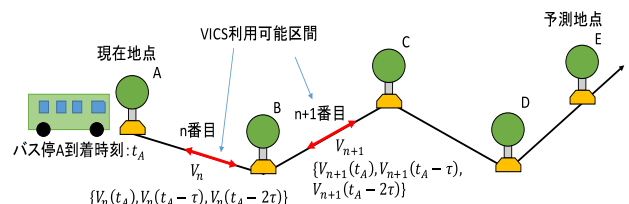


図 4 VICS データの入力値の設定方法

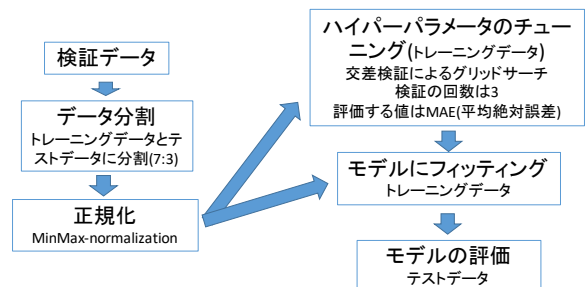


図 5 データの前処理方法

#### 4. CNN の構成と学習に用いるバス時系列データ

始発バス停からの時間と距離のデータを元に、始発バス停から一定距離ごとの平均速度のデータを作成した。それらから得られる予測基準バス停までのデータを入力とし、目標バス停での時刻表からの遅延を出力（目的変数）として CNN による予測を行う。図 6 に入力データ、出力データの関係を示す。

図 7 に検証に使用したモデルのハイパーパラメータ構成を示す。これは 20 パターン程度の構成を試して最も汎化誤差（学習に用いなかったデータを入力した際の誤差）が小さい構成を選択した。学習効率を上げるために Batch Normalization を導入した。また全結合 2 層に対する追加入力として、予測基準バス停における時刻表からの遅延時間を入力した。

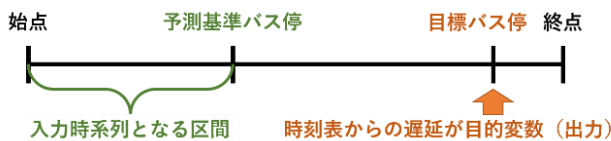


図 6 CNNの入出力データの関係

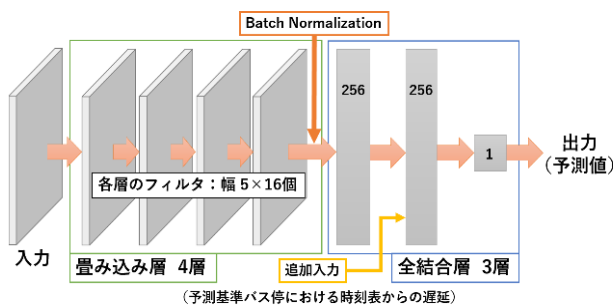


図 7 使用した CNN のモデル構成

#### 5. VICS データを用いた機械学習による検証結果

検証路線は、線形単回帰による予測で紹介した路線 1(鵜沢営業所 07:07 発~甲府駅 08:03 着)である。

図 8 に機械学習モデルごとの各 VICS データ入力方法における的中率の平均値を示す。図 8 は機械学習モデルごとに比較を行うため、各バス停において予測対象地点の到着時刻を予測した際の的中率(Hitting Percentage)を VICS データの入力方法ごとに予測を行なった全地点で合計し平均した図となっている。横軸は機械学習モデル名であり、直近の VICS データのみを利用したケースを one、平均を simple、加重平均を weight、3 つの入力全てをそのまま利用したケースを 3input、VICS データを利用しないケースを none としている。図 9 を見ると、全てのモデルにおいて VICS データを利用しない none の的中率が一番高く、VICS データを一番多く利用する 3input において一番精度が悪いことがわかる。これは、VICS データが予測に悪影響を及ぼし、精度が悪くなったと考えられる。One、simple、weight のモデルにおいては VICS データの入力を各区分において 3 つから 1 つに変換しデータ数を削減している。そのため、悪影響を受けにくく 3input よりも精度の下がり具合は低くなっているが、VICS データを利用しない none に比べ精度向上は見られなかった。VICS データを加えたことによる精度の悪化の原因としては、検証路線が長距離であり、リアルタイムな道路交通状況を入力として加えた

しても実際に VICS 利用区間に到着する時刻と情報を取得する時刻に大きな差が生じているため、交通状況データが不必要なデータとして学習されてしまっていることが考えられる。また、ボトルネック地点を超えてしまえば単回帰のモデルの精度は向上する。つまり、各バス停通過到着時刻と予測対象地点到着時刻に強い相関が生まれる。そこに、VICS データが入力として加わると、相関関係を崩してしまい精度が悪化する。図 9 に Linear モデルの各バス停における的中率の変化を示す。図を見やすくするため、一番精度の悪かった 3input と一番優れていた none、そして入力を 1 つにしたケースにおいて一番精度の良かった one の 3 つを示した。ボトルネック地点の手前の 21~29 番目の停留所に注目すると、none よりも one、3input の方が的中率は高く精度が良くなっていることがわかる。また、ボトルネック付近に近く、ボトルネック地点の VICS データ利用区間との時間差が少ない手前の地点では、精度の向上が見られた。

ボトルネック地点よりも前の地点における各モデルでの的中率の平均を図 10 に示す。Linear モデルにおいては VICS データを含めた場合の方が、none よりも精度が良いことがわかる。しかし、SVR モデルの none が全体としては良い結果であった。図 11 に Linear モデルと SVR モデルの none における各バス停での的中率の変化を示す。SVR モデルでは、ボトルネック前の地点において Linear モデルに比べ精度が良い。これは、SVR モデルでは到着時刻から渋滞の特性を非線形な関係として捉えることができたためであり、このことが SVR モデルの none の精度が一番良くなった理由と考えられる。

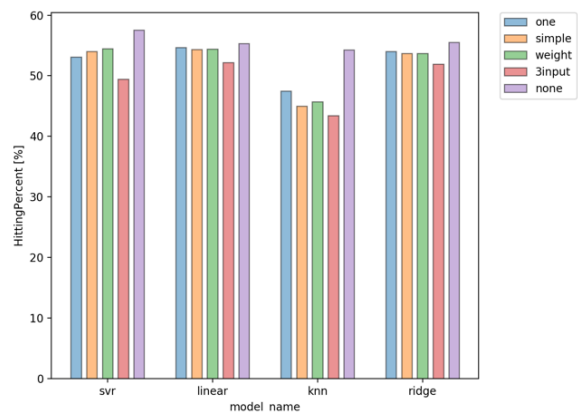


図 8 各機械学習モデルの的中率の平均値

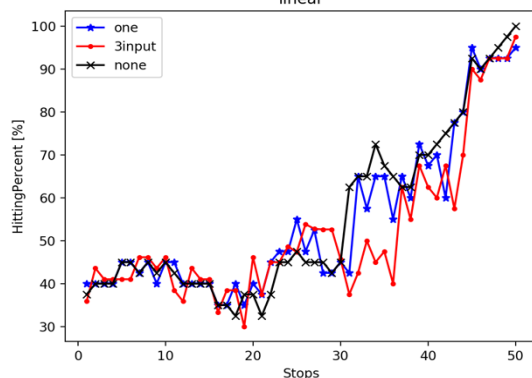


図 9 VICS データ入力方法による的中率の違い (Linear モデル)

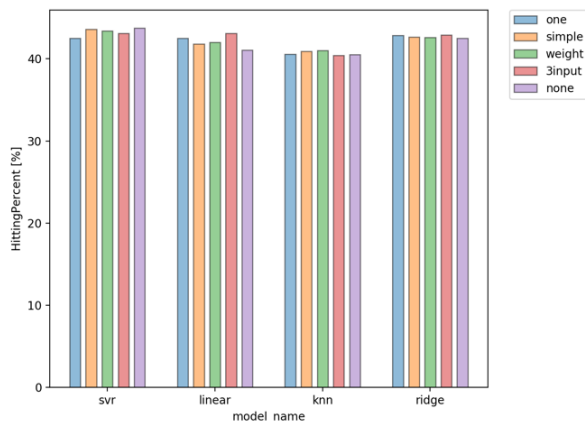


図10 ボトルネック地点より前の停留所における各機械学習モデルの的中率の平均値

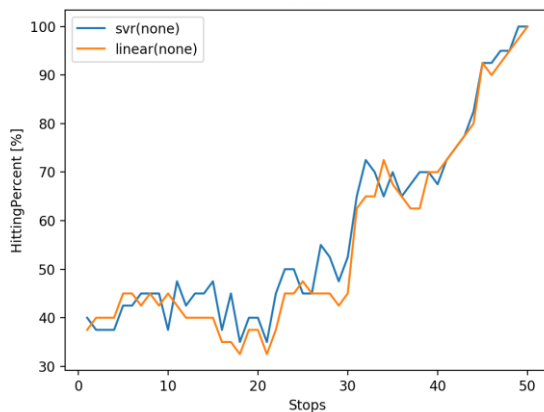


図11 SVRモデルとLinearモデルにおけるVICSデータなしの場合の的中率の差

## 6. CNNとSVR, k-nn, ridge, LRの精度比較

CNNとSVR, k-nn, ridge回帰, LR(予測基準バス停での遅延による予測)について精度比較を行う。データには、2015/01/01から2018/12/31までのものを訓練と検証用に8:2の割合で分割して用いた。基準バス停毎の予測誤差のグラフを図12に示す。路線は単回帰でボトルネックの例に挙げた鵜沢07:07発~甲府駅08:03着線、各バス停を基準とするモデルを作り終点2つ前のバス停での遅延を予測した時のMAE(Mean Absolute Error)のグラフである。ここでMAEは、予測値と実測値の差の単位を秒(s)にしたものとした。ボトルネックのあるバス停(30あたり)からバス停10箇所前あたりまでの精度が他の手法と比べて高いことがわかる。しかし、始発から12バス停あたりまでの精度はほか手法と比べて低いことがわかる。

図13に敷島営業所07:53発~山梨大学附属病院08:44着線での精度比較結果を示す。この路線では中央付近のバス停に軽いボトルネックがある。図12の路線とは違い、始発から中央のバス停付近まで他の手法よりも精度がよく、それ以降は同じような結果となっていることがわかる。

以上の2路線の例から、CNNはボトルネック直前での精度がほか手法に比べ高いことがわかった。しかし、ボトルネック以降に関してはほとんど変わらなかった。

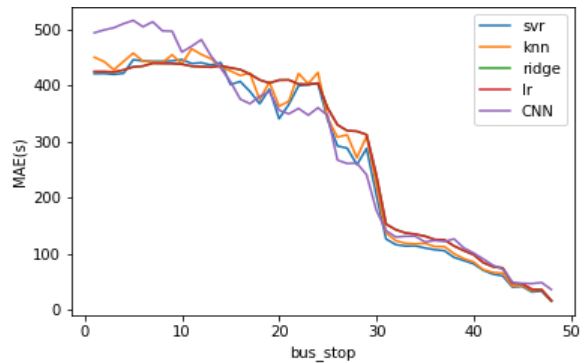


図12 CNNとSVR, k-nn, ridge, LRの精度比較結果

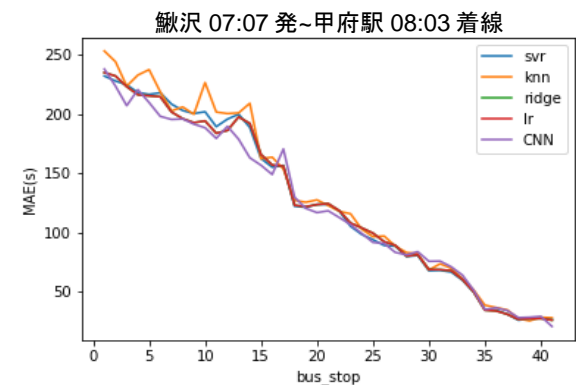


図13 CNNとSVR, k-nn, ridge, LRの精度比較結果

敷島営業所 07:53 発~山梨大学附属病院 08:44 着線

## 7. おわりに

バス到着時刻の予測が難しい地方において、予測精度向上のためCNN, SVR, k-nn, ridge回帰, LRに時系列データを入力し、予測精度を比較した。大きな差はなかったが、ボトルネックとなるバス停直前での精度はCNNが高いことがわかった。また、リアルタイムのデータを入力するためにVICSデータを利用したSVR, k-nn, ridge回帰, LRでの精度比較を行った。文献[6]と同じく、SVRでの予測精度が最も高かった。VICSデータの利用に関しては必ずしも精度を上げるわけではなかった。

### 参考文献

- [1]. やまなしバスコンシェル <http://busmaps.jp/yamanashi/>
- [2]. Mehmet Altinkaya, Metin Zontul, "Urban Bus Arrival Time Prediction: A Review of Computational Models", Int. J. Recent Tech. and Eng. (IJRTE), vol.2, pp.164-169, 2013.
- [3]. 内村圭一, 成松裕介, 江藤旭秀, 胡振程, "バスロケーション情報を用いたバス停間所要時間予測", 国際交通安全学会誌, Vol.32, pp.224-231, 2007.
- [4]. 前川裕一, 林夏美, 牧野友哉, 白石陽, "バス到着時刻予測におけるバス所要時間データと乗降客数データの活用", 「マルチメディア通信と分散処理ワークショップ」, pp.165-171, 2013.
- [5]. 轟朝幸, 川崎智也, 野村大智, 横関敬裕, "ニューラルネットワークを用いた路線バスの遅延時間予測", 第36回交通工学研究発表論文集, pp.489-493, 2016.
- [6]. Bin Yu, William H.K. Lam, Mei Lam Tam, "Bus arrival time prediction at bus stop with multiple routes", Transportation Research Part C, 19, pp.1157-1170, 2011.