Aggregate Learning for Mixed Frequency Data

Takamichi Toda
Daisuke Moriwaki
Kazuhiro Ota
toda_takamichi@cyberagent.co.jp
moriwaki_daisuke@cyberagent.co.jp
ota_kazuhiro@cyberagent.co.jp
CyberAgent, Inc.
Shibuya, Tokyo, Japan

ABSTRACT

Large and acute economic shocks such as the 2007-2009 financial crisis and the current COVID-19 infections rapidly change the economic environment. In such a situation, real-time analysis of regional heterogeneity of economic conditions using *alternative data* is essential. We take advantage of spatio-temporal granularity of alternative data and propose a **Mixed-Frequency Aggregate Learning (MF-AGL)** model that predicts economic indicators for the smaller areas in real-time. We apply the model for the real-world problem; prediction of the number of job applicants which is closely related to the unemployment rates. We find that the proposed model predicts (i) the regional heterogeneity of the labor market condition and (ii) the rapidly changing economic status. The model can be applied to various tasks, especially economic analysis.

CCS CONCEPTS

• **Theory of computation → Semi-supervised learning**; • **Applied computing → Economics**; • **Information systems → Geographic information systems**.

KEYWORDS

Nowcasting, Aggregate Output learning, Aggregate Learning, Semi-supervised learning, Mobility data, GPS, Location data, Unemployment, Mixed-frequency data, Economic policy

1 INTRODUCTION

Devastating external economic shocks such as the 2007-2009 financial crisis and COVID-19 infections rapidly change economic circumstances. Amid these situations, forecasting of current economic status, Nowcasting, is essential. Nowcasting often takes advantage of *alternative data*, non-standard data such as search queries, location data, SNS data, and satellite images [2, 4, 6, 11, 18, 26].

In response to the surging demand for alternative data, tech giants have started providing economic indexes based on their proprietary data [1, 10, 20]. However, their index does not help us understand the whole picture of the economy itself because the relationship between these indices and familiar economic indicators such as GDP is unclear. To fill this gap, many nowcasting/forecasting models that use high-frequency data such as Google search query to predict economic variables are proposed [2–4, 18, 24–26]. Models for mixed frequency data have been actively studied by economists [7, 8].

One caveat of these studies is that they do not fully extract the granularity of the alternative data. Although the predictor variables have a greater granularity, the predicted values are aggregated at the state or even national level. However, governments, especially local governments need to deal with heterogeneity among small areas.

The problem is that the official statistics are not granular. This is a typical *aggregate learning* problem [19]. Recently, the problem has been taking attention from an increasing number of researchers [5, 15].

We, here, combine mixed-frequency data literature in economics and aggregate learning literature in machine learning to fully utilize the richness of alternative data and provide an example of useful application. Our **Mixed-Frequency Aggregate Learning (MF-AGL) model** takes advantage of spatio-temporal granularity of alternative data and predicts economic indicators in smaller areas in real-time, which are not possible by standard forecasting models. The model also updates the prediction in real-time using high-frequency data without high-frequency target data. To train the model, we define a novel loss function for spatio-temporally aggregated label data and granular predicted values. More specifically, we aggregate predicted values for small areas and calculate loss using an aggregated level. We also calculate the loss for each high-frequency feature so that the model can learn the mixed-frequency structure of the data. That means we reuse the same label data over and over.

We apply the model to the real-world problem; prediction of the number of job applicants for smaller areas in Japan. The number of job applicants reflects the condition of the labor market because job applicants are also unemployed persons. An acute increase in the number of job applicants implies deterioration of labor market conditions.

To our best knowledge, the present work is the first to propose a novel method combining mixed frequency data and aggregate learning. We also demonstrate its practical importance in a real-world application. While we applied the model to the nowcasting of the labor market, the model can be applied to any task that contains (i) infrequent and aggregated indices such as GDP and (ii) spatio-temporally granular data.

2 PROBLEM SETTING AND METHOD

2.1 Problem Setting

Let $y_t^p \in \mathbb{R}$ be a *target* variable which is of interest of economists (e.g. GDP). p stands for some larger area such as nation and state and t stands for some longer time period such as quarter and month. Let $x_\tau^q \in \mathbb{R}$ be *feature* variable which is correlated with the target variable y_t^p. q stands for some smaller area such as city and county and τ stands for shorter time period such as day. The difference between t and τ, and p and q are the *granularity*. In particular, an area p is divided by multiple small areas qs and time period t is divided by multiple short time period τ. We use mapping function μ and π to describe the relationship. In particular, $\mu(\tau) = t$ indicates that time τ belongs to t and $\pi(q) = p$ indicates area q belongs to p.

Our goal is to find a predictor f which predict y_{t0}^q from granular data $(x_\tau^q)_{\tau \leq \tau0}$, where $\mu(\tau_0) = t_0$, (i.e. τ_0 belongs to t_0). Notice that the superscript is not larger area p but small area q. That is, f predicts y_t^q instead of y_t^p.

Table 1 illustrates the data structure of our problem. Feature vector $\mathbf{x}$ is observed for City 1 and 2 and for each day. But output value y is only observed for prefecture (assume the prefecture comprises of only two cities) and each month. We want to predict monthly values for each city (i.e., y_{Jan}^1 and y_{Jan}^2). Since feature vector $\mathbf{x}$ is collected in real-time, we want to update the prediction using the latest information. As shown in Table 1 the predicted values are changing according to the feature vector. That is, $\hat{y}_{\text{Jan}}|\mathbf{x}_{\text{Jan1}} \neq \hat{y}_{\text{Jan}}|\mathbf{x}_{\text{Jan2}}$. In this way, we can fully utilize the real-time and granular data for the forecast.

Table 1: The predicted data is updated every day using latest information although the label data is only available for spatially and temporally aggregated value.

	output	input		pred	
	Pref	Pref.A		Pref.A	
		City1	City2	City1	City2
Jan 1	y_{Jan}	$\mathbf{x}_{\text{Jan1}}^1$	$\mathbf{x}_{\text{Jan1}}^1$	$\hat{y}_{\text{Jan}}^1 \mid \mathbf{x}_{\text{Jan1}}^1$	$\hat{y}_{\text{Jan}}^2 \mid \mathbf{x}_{\text{Jan1}}^2$
Jan 2		$\mathbf{x}_{\text{Jan2}}^1$	$\mathbf{x}_{\text{Jan2}}^2$	$\hat{y}_{\text{Jan}}^1 \mid \mathbf{x}_{\text{Jan2}}^1$	$\hat{y}_{\text{Jan}}^2 \mid \mathbf{x}_{\text{Jan2}}^2$
$\cdots$		$\cdots$	$\cdots$	$\cdots$	$\cdots$
Jan 31		$\mathbf{x}_{\text{Jan31}}^1$	$\mathbf{x}_{\text{Jan31}}^2$	$\hat{y}_{\text{Jan}}^1 \mid \mathbf{x}_{\text{Jan31}}^1$	$\hat{y}_{\text{Jan}}^2 \mid \mathbf{x}_{\text{Jan31}}^2$
Feb 1	y_{Feb}	$\mathbf{x}_{\text{Feb1}}^1$	$\mathbf{x}_{\text{Feb1}}^2$	$\hat{y}_{\text{Feb}}^1 \mid \mathbf{x}_{\text{Feb1}}^1$	$\hat{y}_{\text{Feb}}^2 \mid \mathbf{x}_{\text{Feb1}}^2$

While we conduct spatial disaggregation, we directly predict the aggregated value for a longer period. That is, we predict y_t^q rather than y_τ^q. This seems simple supervised learning but it is not. To see this, let τ be Jan 15, 2020. Then t is Jan 2020. We have data only for the past. That is, we have $\mathbf{x}_{\text{Jan1}}^q, \mathbf{x}_{\text{Jan2}}^q, \cdots, \mathbf{x}_{\text{Jan15}}^q$ but do not have $\mathbf{x}_{\text{Jan16}}^q, \mathbf{x}_{\text{Jan17}}^q, \cdots, \mathbf{x}_{\text{Jan31}}^q$. Nowcasting predicts the aggregated

value using incomplete time series data [8]. For economists, the monthly value y_t^q is more informative than the daily value y_τ^q. y_τ^q represents the economic status in a very short term while y_t^q represents the forecast of the longer time period. For example, let y_t^q be the unemployment rate for January 2020 and y_τ^q be that for January 1, 2020. While the daily movement of the unemployment rate is interesting for stock traders who are eager to know the short-term fluctuation, the forecast of the monthly unemployment rate is of prime interest for policymakers who need to know the trend of economic status. Furthermore, policy interventions take some time to be implemented.

On the other hand, geographically granular estimates y_t^q are much more useful than aggregated value especially for local governments who need precise information about their local economy. Nevertheless, our method is easy to transform into a daily version. In sum, our goal is to find a good predictor $f(\mathbf{x}_\tau^q)$ for y_t^q, $\mu(\tau) = t$.

2.2 Aggregate Learning for Mixed-Frequency Data

Following the aggregate learning literature[5, 19], we first define the aggregation function as,

$$\hat{y}_t^p = \sum_{q \in p} \omega_q f(\mathbf{x}_\tau^q, \boldsymbol{\phi}_\tau^q), \tag{1}$$

where $\omega_q \in [0, 1]$ is weight. The weight controls for the share of the values for each granular area in the large area. In the simplest case (including the application discussed below), ω_q is set to one. In these cases, f can learn the actual values of the target from the features. However, when we only obtain *normalized* features such as population per acre or averaged age, we need a weighted sum. In these cases, the weights are pre-determined based on real data and knowledge. Area and population are often surveyed at a fine granularity by censuses and can be used as good proxies of the share.

2.2.1 Mixed-Frequency Aggregate Learning Model. Predictor f predicts the outcome values for small areas. In contrast to standard supervised learning, true values for the predictor can not be observed. In another word, the predictor predicts *latent* values. The main feature vector $\mathbf{x}_\tau^q$ is of granular information such as search queries, posts in social networking service (SNS), point of sales (POS), credit card, and mobility data. The vector $\mathbf{x}_\tau^q$ contains the current value x_τ^q and its lagged values. That is, $\mathbf{x}_\tau^q = (x_\tau^q, x_{\tau-1}^q, x_{\tau-2}^q, \cdots)$. Forecasters want to update the prediction when new data arrive. That is, τ can be any timing. For example, Let t be April 2020 and $\tau_1, \tau_2, \cdots, \tau_{30}$ be April 1st, 2nd, $\cdots$, 30th. Then, $\mathbf{x}_{\tau1}$ and $\mathbf{x}_{\tau2}$ should give different prediction. However, the label data are only available for complete data $\mathbf{x}_{\tau30}^q = (x_{\tau30}^q, x_{\tau29}^q, \cdots, x_{\tau1}^q)$. $x_{\tau30}^q$ is missing in $\mathbf{x}_{\tau29}^q$. In economics, missing data is dealt with by either (i) a training model on data with the same missing structure or (ii) an imputation of missing data. In the above example, a forecaster who adopts (i) uses only data from the 1st to 29th day for each month to keep the same missing structure when training their model. Naturally speaking, it causes a huge loss of information. A forecaster who resorts to (ii) needs to prepare another model to conduct imputation.

In our model, the missing-ness is treated by introducing auxiliary feature vector $\boldsymbol{\phi}$ and non-linearity in the parameters. The vector $\boldsymbol{\phi}_\tau^q$

contains one-hot encoded year, month, day, larger area p, and small area q. By doing so, predictor f can learn the missing structure of the data and appropriately use the information. The non-linearity of the parameter is essential in this sense. We adopt a simple recurrent neural network (RNN) to make the predictor flexible. As a result, we can fully utilize the information and make it end-to-end.

2.2.2   Model Training. The predictor is trained by minimizing the MSE loss,

$$\mathcal{L}(f) = \sum_p \sum_t \sum_{\tau:\mu(\tau)=t} \left(y_t^p - \sum_{q:\pi(q)=p} \omega_q f(\mathbf{x}_\tau^q, \boldsymbol{\phi}_\tau^q) \right)^2. \qquad (2)$$

At each data point (τ, p), the predictor f predicts the values for small area q at τ. The predicted values are weighted average and the loss is calculated.

2.2.3   Prediction. Trained predictor f is used for the prediction using granular data. Forecaster can use real-time data $\mathbf{x}_\tau^q$ to predict the smaller area's current status as $\hat{y}_t^q = f(\mathbf{x}_\tau^q)$. The good news is that the forecaster does not need to re-train the model until new label data y is released. She can re-use the same model for an extended period.

3   EXPERIMENT WITH JAPANESE JOB APPLICATION DATA

We apply the model to predict the number of job applicants in Japan. The Japanese government releases official statistics on the number of persons who file job applications to public employment offices on monthly basis. Job application is required to take up the unemployment insurance benefits. The number of job applications is a good proxy for the number of unemployed persons as the unemployment insurance claims statistics in the U.S. is considered one of the most important economic indicators by economists.

The reasons we chose the problem are the following. First, unemployment is a huge tragedy; It leads to loss of income and also loss of contact with society, which causes economic and mental hardship. Real-time analysis is essential for a swift policy response. Second, the prefecture-level data provided by the reports are too rough for the appropriate policy response. In Japan, each prefecture has a population of several million to ten million. More granular statistics are needed for careful policy intervention. Third, monthly updates of the reports are too infrequent. Looking at monthly data for one or two months ago is not very meaningful especially amid COVID-19.

Fortunately, there is a good *alternative data* for the number of job applicants. As Moriwaki [18] shows GPS data from smartphones has good predictive power for the number of unemployed persons. In this study, we utilize similar datasets. In contrast to [18] who only count the number of GPS readings inside the radius of the offices, we extract rich features from these data.

The whole process is summarized in Fig. 1 We first extract feature vector $\mathbf{z}$ from raw GPS readings taken from mobile apps. Then the visit predictor predicts $\mathbf{x}_\tau^q$, the number of visits to public unemployment offices located in each city. Then another predictor f predicts y_t^q, the number of job applicants for each office. The visit predictor is trained on the different domains and transferred to the task. The transfer learning is explained in Section 3.2.
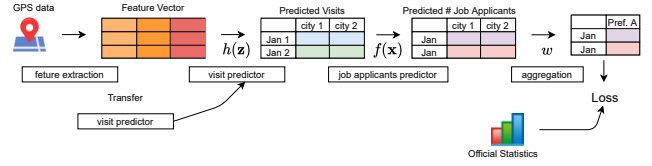

Figure 1: Overview of the data pipeline.

3.1   Datasets

We use three datasets as follows. **Reports on Employment Service** is a monthly official statistic released by the Ministry of Health, Labour and Welfare. The reports contain the monthly number of job applicants by prefecture. The data is publicly available on the webpage of the ministry [17]. **GPS data from Smartphones** is taken from various mobile apps from Jan 1, 2016, to Oct 31, 2020. The data is anonymized and not related to privacy. The number of users ranges from several hundred thousand to several million. The data contains tuples (latitude, longitude, timestamp). The data contains no private information. **Locations of Public Employment Offices** is publicly available on the webpage of the MHLW. We use the location with the GPS trajectories to extract feature values [16].

3.2   Visits Prediction using Transfer Learning

3.2.1   Transfer of Visit Predictor. Since there is no *true* label for public employment office visitors, we transfer the visit predictor trained on the different proprietary source and transfer to the visit prediction for the main task; visits to public employment offices. The predictor is implemented using LightGBM [12].

3.2.2   Feature Extraction from GPS Readings. The crucial challenge is to accurately count visitors from noisy GPS readings. Visited point extraction methods typically rely on (i) the number of stay points, (ii) the stop location, (iii) stop duration, and (iv) speed [9, 13, 21, 22].

3.2.3   Visits Prediction for Public Employment Offices. Unemployed persons visit the nearest public employment office to file an unemployment insurance claim. There are 544 offices in Japan. We divide the entire country into 544 regions based on the location of the offices assuming each region is covered by the nearest office. We use the transferred predictor to predict the visit count to each office.

3.3   Job Applicants Predictor and Aggregation Function

With the predicted visits count, we train the job applicants predictor that predicts the number of job applicants for each public employment office. The predictor uses past visit count and dummy variables extracted from the date of making a prediction.

The training/prediction model is a simple recurrent neural network. This model uses an LSTM layer and a multi-layer perceptron to predict a daily visit count of small areas, and each prediction is aggregated to predict a monthly visit count of large areas according to Eq. (1). We use Adam [14] solver for optimization with

$\beta 1 = 0.9, \beta 2 = 0.999$, initial learning rate = 0.0001, no weight decay and no learning rate decay. We train our models for a total of 600 epochs with a batch size of 1 and the MSE loss described in Eq. (2) on Tesla V100 GPU. We implemented the whole network in PyTorch [23].

3.4 Spatial Disaggregation using MF-AGL model

Fig. 2 shows the four maps represent regional heterogeneity of the change of the number of job applicants. The maps are generated by data from the actual Reports on Employment Service for October 2020 (Ground Truth), predictions by the proposed model (MF-AGL), predictions by an auto-regressive model (AR), and predictions by a Random Forest model (RF). The AR model uses the number of job applications of past 11 months (i.e., 11 lags) as inputs, i.e., $\hat{y}_t^p = f^{AR}(y_{t-1}^p, \cdots, y_{t-11}^p)$. RF model uses dummy variables for a year, month, and prefecture as well as the number of job applications of the past 11 months (i.e., 11 lags).
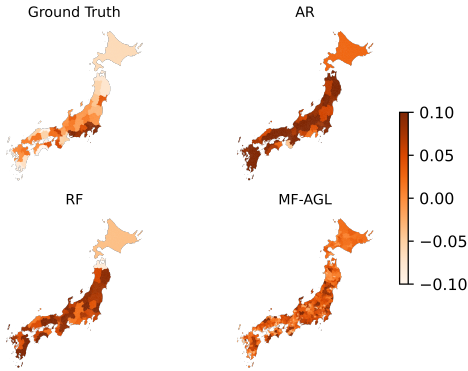

Figure 2: Year-over-year change in the number of job applicants in Japan. Good (lighter)-bad (darker) conditions of labor market. Ground Truth shows the actual data from official statistics;AR shows the prediction made by the AR model with 11 lags;RF shows the predictions made by Random Forest model with 11 lags and year, month, and prefecture dummy as input;MF-AGL shows the prediction made by MF-AGL model. MF-AGL model shows many granular predictions.

The MF-AGL model is trained on the data from October 2016 to September 2020. Then the model use feature extracted on October 31, 2020. The AR and RF model is trained on the data from October 2016 to September 2020 and uses the data from September 2020 as input. It seems unfair that only the MF-AGL model uses the data from October 2020. However, this is the strength of the nowcasting model.

Now let's turn to the maps in Fig. 2. One significant difference of the MF-AGL model is its geographical granularity. While the other three figures only tell that the southeast areas are in bad condition (darker), the MF-AGL model shows that there is a mix of bad and good conditions at a granular level. We can see there are darker areas in the granular map that are bright in the ground truth data. The local governments need to take care of these hidden problems.

The other finding is that the AR and RF models are not good at prediction. To see this, we first aggregate the prediction by the MF-AGL model at prefecture level and calculate the Mean Absolute Percentage Error (MAPE), a standard metric to evaluate forecasting models.

The results are shown in Table 2. The MF-AGL obtained the lowest error. Although the prediction performance at the aggregate level is not the priority of the MF-AGL model, this result highlights the robustness of the prediction of the model.

Table 2: Mean Percentage Errors by models

model	mape (%)	Std. Err	N
AR	13.41	(0.66)	47
RF	8.11	(0.67)	47
MF-AGL	**7.78**	(1.03)	47

Note: The mean percentage error by each model. The values are calculated for the prediction shown in Fig. 2.

3.5 Real-time Labor Market Analysis using MF-AGL Model

Finally, we demonstrate another practical usefulness of our model. Fig. 3 shows how the model updates the prediction using the real-time data. The model shows the year over year for each city and for each day. That is, we first *predict* the disaggregate number of job applicants on the same day of the previous year and calculate the year-over-year change in the number of job applicants. The figure implies the rapid improvement in the labor market during October 2020. The possible reason is the peak-out of the second wave of the COVID-19 pandemic. The number of cases was dramatically decreased in September and stable in October.
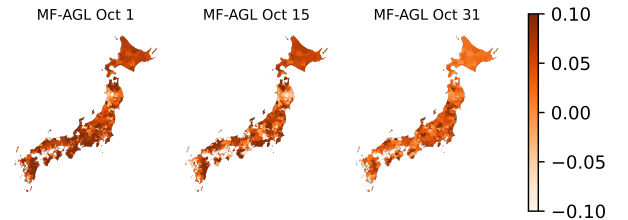

Figure 3: Changes in the number of job applicants in Japan. The prediction is done by MF-AGL model uses the input available at each day.

4 CONCLUSION

In this work, we proposed a novel aggregate learning method that adapts to the mixed-frequency data. The model predicts spatio-temporal changes in the economic indices without granular label data. We proposed an LSTM-based simple architecture and the loss function for the training.

We applied the model to a real-world task that is the prediction of the number of job applicants in Japan. Our MF-AGL model well

predicts the regional heterogeneity at the sub-prefecture level and also the rapid change in the labor market condition during a month.

The present model can be applied to broad areas including GDP prediction, labor market prediction, and industrial production prediction. The direction of the future work can be the extension to the other domain and more improvement of the model using more fine-grained architectures.

ACKNOWLEDGMENTS

REFERENCES

[1] Apple Inc. 2020. COVID-19 - Mobility Trends Reports - Apple. https://webcache.googleusercontent.com/search?q=cache:0e-S9-MxVDYJ:https://covid19.apple.com/mobility+&cd=1&hl=ja&ct=clnk&gl=jp.
[2] Nikos Askitas and Klaus F. Zimmermann. 2009. Google Econometrics and Unemployment Forecasting. (2009).
[3] Hyunyoung Choi and Hal Varian. 2012. Predicting the Present with Google Trends. *Economic Record* 88, s1 (2012), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x
[4] Francesco D'Amuri and Juri Marcucci. 2010. 'Google It!' Forecasting the US Unemployment Rate with A Google Job Search Index. *SSRN Electronic Journal* (2010). https://doi.org/10.2139/ssrn.1594132
[5] Guillaume Derval, Frédéric Docquier, and Pierre Schaus. 2020. An Aggregate Learning Approach for Interpretable Semi-Supervised Population Prediction and Disaggregation Using Ancillary Data. In *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science)*, Ulf Brefeld, Elisa Fromont, Andreas Hotho, Arno Knobbe, Marloes Maathuis, and Céline Robardet (Eds.). Springer International Publishing, Cham, 672–687. https://doi.org/10.1007/978-3-030-46133-1_40
[6] Jaqueson K. Galimberti. 2020. Forecasting GDP Growth from Outer Space. *Oxford Bulletin of Economics and Statistics* 82, 4 (2020), 697–722. https://doi.org/10.1111/obes.12361 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12361.
[7] Eric Ghysels. 2016. Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics* 193, 2 (Aug. 2016), 294–314. https://doi.org/10.1016/j.jeconom.2016.04.008
[8] Eric Ghysels, Arthur Sinko, and Rossen Valkanov. 2007. MIDAS Regressions: Further Results and New Directions. 26, 1 (2007), 53–90. https://doi.org/10.1080/07474930600972467
[9] Lei Gong, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. *Procedia - Social and Behavioral Sciences* 138 (July 2014), 557–565. https://doi.org/10.1016/j.sbspro.2014.07.239
[10] Google Inc. 2020. COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/?hl=en.
[11] Agustín Indaco. 2020. From twitter to GDP: Estimating economic activity from social media. *Regional Science and Urban Economics* 85 (Nov. 2020), 103591. https://doi.org/10.1016/j.regsciurbeco.2020.103591
[12] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2017-12-04) *(NIPS'17)*. Curran Associates Inc., 3149–3157.
[13] Ilkcan Keles, Matthias Schubert, Peer Kröger, Simonas Šaltenis, and Christian S. Jensen. 2017. Extracting visited points of interest from vehicle trajectories. In *Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data (GeoRich '17)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3080546.3080552
[14] Diederik Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*. 11.
[15] Ho Chung Leon Law, Dino Sejdinovic, Ewan Cameron, Tim CD Lucas, Seth Flaxman, Katherine Battle, and Kenji Fukumizu. 2018. *Variational Learning on Aggregate Outputs with Gaussian Processes*. arXiv:1805.08463 [cs, stat] http://arxiv.org/abs/1805.08463
[16] and Labour Ministry of Health, Welfare. [n.d.]. Location of Hellowork. https://www.mhlw.go.jp/kyujin/hwmap.html
[17] Ministry of Health, Welfare, and Labour. [n.d.]. Reports on Employment Service. https://www.mhlw.go.jp/toukei/list/114-1b.html
[18] Daisuke Moriwaki. 2020. Nowcasting Unemployment Rates with Smartphone GPS Data. In *Multiple-Aspect Analysis of Semantic Trajectories (Lecture Notes in Computer Science)*, Konstantinos Tserpes, Chiara Renso, and Stan Matwin (Eds.). Springer International Publishing, Cham, 21–33. https://doi.org/10.1007/978-3-030-38081-6_3
[19] David R. Musicant, Janara M. Christensen, and Jamie F. Olson. 2007. Supervised Learning by Training on Aggregate Outputs. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, Omaha, NE, USA, 252–261. https://doi.org/10.1109/ICDM.2007.50
[20] COVID-19 Mobility Data Network. 2020. Facebook Data for Good Mobility Dashboard. https://www.covid19mobility.org/dashboards/facebook-data-for-good/.
[21] Kyosuke Nishida, Hiroyuki Toda, and Yoshimasa Koike. 2017. Extracting Arbitrary-shaped Stay Regions from Geospatial Trajectories with Outliers and Missing Points. In *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Computational Transportation Science - IWCTS'15*. ACM Press, Seattle, WA, USA, 1–6. https://doi.org/10.1145/2834882.2834884
[22] Kyosuke Nishida, Hiroyuki Toda, Takeshi Kurashima, and Yoshihiko Suhara. 2014. Probabilistic identification of visited point-of-interest for personalized automatic check-in. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*. ACM Press, Seattle, Washington, 631–642. https://doi.org/10.1145/2632048.2632092
[23] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
[24] Jaroslav Pavlicek and Ladislav Kristoufek. 2015. Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries. *PLOS ONE* 10, 5 (May 2015), e0127084. https://doi.org/10.1371/journal.pone.0127084
[25] Tanya Suhoy. 2008. Query Indices and a 2008 Downturn: Israeli Data. (2008), 34.
[26] Hal R. Varian and Hyunyoung Choi. 2009. *Predicting the Present with Google Trends*. SSRN Scholarly Paper ID 1659302. Social Science Research Network, Rochester, NY.