COVID-EENet: Predicting Fine-Grained Impact of COVID-19 on Local Economies

Doyoung Kim[1], Hyangsuk Min[1], Youngeun Nam[1], Hwanjun Song[2], Susik Yoon[3], Minseok Kim[1], Jae-Gil Lee[1]*

[1]KAIST, [2]NAVER AI Lab, [3]University of Illinois Urbana-Champaign
{dodokim, hyangsuk.min, youngeun.nam, minseokkim, jaegil}@kaist.ac.kr,
hwanjun.song@navercorp.com, susik@illinois.edu

Abstract

Assessing the impact of the COVID-19 crisis on economies is fundamental to tailor the responses of the governments to recover from the crisis. In this paper, we present a novel approach to assessing the economic impact with a large-scale credit card transaction dataset at a fine granularity. For this purpose, we develop a *fine-grained* economic-epidemiological modeling framework ***COVID-EENet***, which is featured with a two-level deep neural network. In support of the fine-grained EEM, *COVID-EENet* learns the impact of nearby mass infection cases on the changes of local economies in each district. Through the experiments using the nationwide dataset, given a set of active mass infection cases, *COVID-EENet* is shown to precisely predict the sales changes in two or four weeks for each district and business category. Therefore, policymakers can be informed of the predictive impact to put in the most effective mitigation measures. Overall, we believe that our work opens a new perspective of using financial data to recover from the economic crisis. For public use in this urgent problem, we release the source code at https://bit.ly/covideenet.

Introduction

Motivation

The COVID-19 pandemic is far more than a health crisis. It has almost paralyzed economic activity, as countries impose strict restrictions on moves to contain the virus. In most countries, countless small businesses have collapsed with unemployment rising by about 42% and per capita income falling by around 7% (Acs and Karpman 2020; Coibion, Gorodnichenko, and Weber 2020; Sumner et al. 2020). The economic consequences of COVID-19 represent the largest economic shock that the world has experienced in decades. The crisis highlights the need for urgent measures to mitigate the economic shock of the epidemic, protect the vulnerable population, and pave the way for continued recovery (Singh 2020). Therefore, the immediate priority of policymakers has been to curb economic damage (Neumann-Böhme et al. 2020).

To efficiently manage the economic crisis caused by the pandemic with limited resources, the policy-making process should be evidence-informed. That is, a good supply of quality evidence and positive prospects should support the policy. Although COVID-19 vaccines have been developed, research into the economic crisis to answer the questions that policymakers deliberate over has not been fully conducted. In lieu of engaging in deliberation about the issue, policy responses have simply relied upon economic statistics such as the gross domestic product (GDP), unemployment rate, and demographics (Hale et al. 2020; Elgin, Basbug, and Yalaman 2020; Cheng et al. 2020).

However, policy response based on this *coarse-grained* economic consequence faces the fundamental limitation that *local* economies have been subject to various degree of impact by the COVID-19 recession and showed unequal recovery trajectories (Demirguc-Kunt, Lokshin, and Torre 2020). That is, the pandemic hit areas and business sectors differently, e.g., higher impact on clubs than on restaurants (Harris 2020; Callinan and MacLean 2020; Sidhu et al. 2020). Without a full understanding of the *fine-grained* economic impact of COVID-19, policy response could be uncertain and inappropriate. For example, many countries provide the economic impact payments to damaged economic sectors, but deciding the recipient and amount may not reflect the real loss by the pandemic (Rahman et al. 2020). Thus, the impact of COVID-19 on local economies should be examined at a finer granularity based on economic activity big data.

Research Problem and Goal

We present *fine-grained economic-epidemiological modeling (EEM)* developed with close collaboration with a major credit card company in South Korea. We call our deep neural network (DNN)-based EEM framework ***COVID-EENet***. In support of fine-grained EEM, as shown in Figure 1, *COVID-EENet* is capable of modeling the impact of each *mass infection case* on the amount of economic activities in consideration of the *economy*, *geography*, and *epidemic* aspects. That is, *COVID-EENet* learns the various degree of the impact from the COVID-19 outbreak and reveals the factors for the different impact. As a result, given recently occurred mass infection cases, *COVID-EENet* can accurately *predict the changes of economic activities (i.e., daily sales) caused by the mass infection cases* per business category and district in the near future (e.g., two or four weeks).

Accordingly, for high-level insights, *COVID-EENet* en-

*Jae-Gil Lee is the corresponding author.

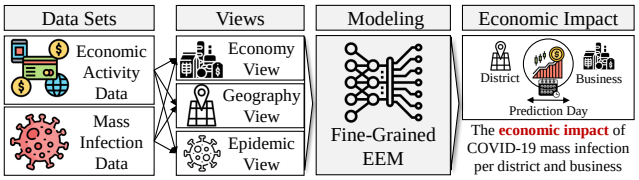

Figure 1: The flow of the fine-grained EEM by *COVID-EENet*, which is capable of modeling the impact of mass infection on economic activities.

ables us to comprehensively answer two questions: (i) which local economies are the most vulnerable? and (ii) what kinds of disparities determine the local economies vulnerable to COVID-19? Answering these questions provides policymakers with motivations or quality evidence for various policies ranging from stimulus measures to government campaigns, as well as strategic ways to practically implement policies. For example, because the framework reveals the business-geography-epidemic disparities in economic damage, the finding can lead to economic stability by motivating selective and proactive funding policies for local economies that are expected to severely suffer from economic damage.

The fine-grained EEM in *COVID-EENet* is realized by virtue of a large-scale, fine-grained economic activity dataset, more specifically, an *aggregated* credit card transaction dataset, which is provided by our collaborator[1]. The dataset contains daily sales for each business category and each district in South Korea for the years from 2019 to 2020. All registered offline and online stores are classified into 34 business categories. The total number of records (daily sales per business category and district) even exceeds 408 million. Overall, the dataset represents a large body of economic activities at a fine granularity—for each combination of 34 business categories, 183 districts, and 730 days. In addition, we collected 150 mass infection cases that occurred in Seoul from February to December 2020.

In order to precisely find complex patterns from the economic-epidemiological dataset, *COVID-EENet* consists of a microscopic encoder and a macroscopic aggregator. First, a *microscopic encoder* models the impact of a *specific* mass infection case on a target district, considering the business category and district where the mass infection case occurred as well as the severity trend of that mass infection case. This encoder involves *multi-view* modeling to combine various influencing factors such as economic similarity, geographic distance, and mass infection size. Second, because multiple mass infection cases affect the target district *simultaneously*, the *macroscopic aggregator* combines the effects of multiple influential mass infection cases. This aggregator exploits a gating mechanism to find the contributions of individual mass infection cases.

Insight Briefs and Contributions

Beyond achieving high predictive power of *COVID-EENet* on the changes of economic activities, due to a nationwide, large-scale credit card transaction dataset, we obtain generalizable insights regarding the economic impact of COVID-19, to list:

- The degree of requiring *person-to-person contact* in businesses is positively correlated with the economic damage.
- Economic activities are confined to their residential areas by severe mass infection cases, so the local businesses in *commercial* districts are more vulnerable.
- The economic impact depends on the *recency*, *severity*, *geographic adjacency*, and *business type* of concurrent mass infection cases.

Overall, the main contributions are as follows:

- We formulate the problem of *fine-grained EEM* for COVID-19, in order to predict the economic activity changes caused by simultaneous COVID-19 mass infection cases. To the best of our knowledge, this is the *first* work that addresses the fine-grained EEM for COVID-19.
- We propose a novel DNN-based EEM framework, ***COVID-EENet***, which consists of a microscopic encoder and a macroscopic aggregator to model respectively the individual and overall effects of mass infection cases.
- We conduct an in-depth analysis for the fine-grained EEM using a nationwide, large-scale credit card transaction dataset. As for the prediction accuracy, *COVID-EENet* outperforms four baseline models by 9.3% and 15.1% on average in terms of the RMSE and the MAE, respectively.
- We provide high-level insights on the economic consequence of COVID-19, which are expected to be very helpful to enact more equitable and effective policies.

	Attribute	Description	Example
	Date	Card transaction date (YYYYMMDD)	20190101
Store	Address	Store district	Gangnam
	Business kind	Store business	Living
Customer group	Nationality	Korean or foreigner	Korean
	Gender	Male, female, or others	Male
	Age	Age groups in 10s	30s
	Household type	Classification by family number, etc.	Single
Sale	Price	Total price of the sales	100,000
	Count	Total count of the sales	10

Table 1: Key attributes in economic activity data.

Economic-Epidemiological Data

Dataset Description

Our economic-epidemiological dataset consists of economic activity and mass infection information. Furthermore, to ascertain the underlying claims of this study, we provide the results from exploratory data analysis in Section A of the supplementary material.

Economic Activity The economic activity dataset contains *aggregated* daily sales, which were paid by a credit card of the data provider (BC Card), for each combination of districts and business categories in South Korea from 2019 to 2020. It is known that approximately 75% of the economically active population in South Korea owns BC credit cards. Table 1 shows the key attributes of the dataset, which can be categorized into the (i) date, (ii) store, (iii) customer group, and (iv) sale information. That is, each record represents the

[1]BC Card (https://www.bccard.com/) is the biggest payment provider in South Korea and has over 38 million customers.

Attribute	Description	Example
Origin place	District the mass infection occurred	Guro
Business category	Business category the mass infection occurred	Service
Title	Known title of the mass infection	Call center
Start date	Date the mass infection started (YYYYMMDD)	20200308
End date	Date the mass infection ended (YYYYMMDD)	20200411
Confirmed cases	Daily number of confirmed cases	$\{100, 80, \ldots, \}$

Table 2: Key attributes in the mass infection data.

sales amount at the stores in the district for the business category by the customers of the customer group on that day. This dataset is *free from privacy concerns* because it does not include any personal identity information. Nevertheless, it is sufficiently fine-grained, where the *daily* sales information is aggregated for each of 183 (districts) $\times$ 34 (business categories) pairs. Moreover, the sales information is broken down into each *customer group*, sharing the nationality, gender, age, household type, and address. Besides, the long period from 2019 to 2020 enables us to compare the economic activities *before and after* the COVID-19 pandemic. While similar credit card datasets have been mainly used for market analysis (Di Clemente et al. 2018), we explore a new perspective of using this dataset to overcome the economic recession by COVID-19.

Mass Infection Because mass infection cases have mostly occurred in populated cities, we focused on those occurred in South Korea's capital, Seoul, whose population is approximately 10 million. For this purpose, we collected the mass infection cases reported by the Seoul Metropolitan Government[2]. Table 2 shows the key attributes of the dataset. The total number of mass infection cases is 150, spanning from February to December in 2020. The criterion for classifying mass infection is orthogonal to our study, and we followed the classification by the Korea Disease Control and Prevention Agency (KDCA)[3]. Each mass infection case typically resulted in over 100 confirmed cases.

Feature Engineering

To generate the input of the proposed EEM from the aforementioned dataset, we extract the features that represent diverse perspectives (views): *economy-view*, *geography-view*, and *epidemic-view*. The economy-view and geography-view features are derived from the credit card dataset in Table 1, and the epidemic-view feature is derived from the mass infection dataset in Table 2. These three views are comprehensively integrated in *COVID-EENet*.

Economy-View Feature The *economy-view* feature is to represent a district (and together with a business category) from the perspective of the consumer economy. Basically, it represents how purchases are *normally* made by the consumers in a district. Thus, to consider normal activities, the consumer activities before the pandemic (i.e., in 2019) are used for extracting this feature, where the economic activities (i.e., sales) are broken down into (i) business and (ii) consumer categories, as follows:

(i) The *business structure* of a district is a probability vector whose dimensionality corresponds to the total number of business categories (34 in this study):
$\langle \ldots, \textit{fraction of the sales for the i-th business category}, \ldots \rangle$.
Here, a fraction is the average of the two fractions separately calculated using price and count as in Table 1.

(ii) The *consumer structure* of a district-business pair is a probability vector whose dimensionality corresponds to the total number of consumer categories (27 in this study[4]):
$\langle \ldots, \textit{fraction of the sales by the j-th consumer category}, \ldots \rangle$.
Again, each fraction is the average of the two fractions created in terms of price and count.

Geography-View Feature The *geography-view* feature is to represent the relationship between *two* districts on the perspective of the physical and social geography, and thus consists of *two* numeric numbers in $[0, 1]$ that correspond to the (i) physical and (ii) social distances, as follows:

(i) The *physical distance* between two districts is simply the normalized Euclidean distance between the borough offices of the two districts. This physical distance is symmetric and invariable to time.

(ii) The *social distance* from a district to another district indicates the amount of personal flow from the former to the latter. More specifically, it represents the fraction of the sales count by the residents of the former among the total sales count in the latter. This social distance is asymmetric and calculated using the consumer activities in 2019 to reflect normal situations.

Epidemic-View Feature The *epidemic-view* feature is to represent the severity trend of a mass infection case along the temporal dimension. More specifically, for each mass infection case, it is a sequence of *quadruples*, $\{\ldots, \langle$ *number of confirmed cases (i) in a specific day, (ii) within a week, (iii) until the specific day, (iv) number of elapsed days since the beginning of the mass infection* $\rangle, \ldots\}$. The length of the sequence is determined by the duration of the corresponding mass infection case.

Problem Formulation

Given a set $\mathcal{D}$ of districts and a set $\mathcal{B}$ of business categories, a *district-business* pair in Definition 1 is used for the target (granularity) of the predictive analysis.

Definition 1. (DISTRICT-BUSINESS PAIR) A *district-business pair* is $(d, b) \in \mathcal{D} \times \mathcal{B}$, where $d \in \mathcal{D}$ and $b \in \mathcal{B}$. □

A set $\mathcal{M}$ of mass infection cases is given as the source of the economic impact, where each *mass infection case* $m \in \mathcal{M}$ is specified by Definition 2.

Definition 2. (MASS INFECTION CASE) A *mass infection case* $m \in \mathcal{M}$ consists of (i) the district-business pair where it occurred and (ii) its epidemic-view feature indicating the number of patients from the mass infection case. □

The *economic impact* on a given district-business pair (d, b) is measured by the change of the sales for the district-business pair compared with the sales made *one year ago*, precisely speaking, 364 days ago to keep the same day of the week. For example, May 4 (Saturday), 2019 is referenced for

[2] https://www.seoul.go.kr/

[3] http://www.kdca.go.kr/

[4] 3 (gender) $\times$ 3 (age) $\times$ 3 (household) $= 27$.

May 2 (Saturday), 2020. Besides, the sales amount is quantified by the "sale price" attribute in Table 1. Then, to predict the economic impact for the future, the *economic impact trend* in Definition 3, which is a sequence of the changes for upcoming w days, is used as the target variable.

Definition 3. (ECONOMIC IMPACT TREND) The *economic impact trend* on (d, b) is $\mathbf{y}_{(d,b)} = \{y(t)\}_{t=1}^{w}$, where

$$y(t) = \frac{\text{sales amount on day } t - \text{sales amount a year ago from day } t}{\text{sales amount a year ago from day } t} \quad (1)$$

and day t is the date after t days from the current date. □

Problem Definition: Finally, the *fine-grained EEM* is formulated by Definition 4.

Definition 4. (FINE-GRAINED EEM) The *fine-grained EEM* is, given a set $\mathcal{D}$ of districts, a set $\mathcal{B}$ of business categories, and a set $\mathcal{M}$ of mass infection cases, to predict the economic impact trend $\mathbf{y}_{(d,b)}$ in Definition 3 for each district-business pair $(d, b) \in \mathcal{D} \times \mathcal{B}$ for upcoming w days. □

Note that the *previous* economic impact trend is *not* given as the input of the problem for practical usability in real-world scenarios, where the credit card transactions do not become available immediately. Hence, the problem is much more challenging than a typical time-series prediction problem. Once a trained model is ready, the goal is to predict the economic impact trend *only* using the information about mass infection cases at the current date.

Methodology: *COVID-EENet*

Overview

Figure 2 illustrates the two-level architecture of *COVID-EENet*. A *microscopic encoder* learns a hidden representation that represents the economic impact of a given mass infection case m on a target district d. This encoder comprises the economy-view sub-encoder, the geography-view sub-encoder, and the epidemic-view sub-encoder, each of which is responsible for the corresponding view features. These sub-encoders respectively produce the economy-view representation (*ECR*), the geography-view representation (*GER*), and the epidemic-view representation (*EPR*), which are merged into the microscopic representation (*MIR*) by the view combiner, as shown in Figure 2(a). Then, the *macroscopic aggregator* combines the economic impact of each mass infection case m into the macroscopic representation (*MAR*), followed by the gating module that determines how much each mass infection case affects the district-business pairs of a target district d, thereby predicting the economic impact trend, as shown in Figure 2(b). See Section B of the supplementary material for the pseudocode of its training algorithm.

Phase 1: Microscopic Encoder

Economy-View Sub-Encoder An outbreak district has more economic impact on other districts if they have higher *structural similarity* in terms of *business* or *consumer* distributions. Further, an outbreak business has more economic impact on a certain business with higher *business similarity* in other districts. To this end, the economy-view sub-encoder transforms the economy-view feature into a business-structure similarity, a consumer-structure similarity, and an outbreak-business similarity, and then combines them to generate the economy-view representation.

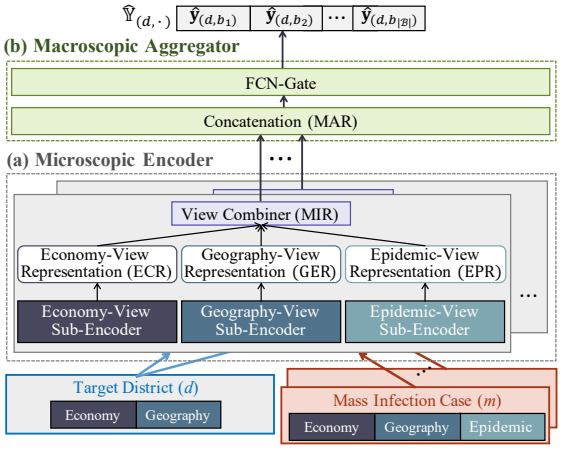

Figure 2: The two-level architecture of *COVID-EENet*.

For a district-business pair (d, b), we use a *district-business embedding* of Definition 5 in a newly defined embedding space to get a higher learning capability in calculating each similarity.

Definition 5. (DISTRICT-BUSINESS EMBEDDING) A *district-business embedding* $e_b \in \mathbb{R}^n$ is an n-dimensional vector for a district-business pair (d, b). A set of district-business embedding vectors in a district d forms a *district-business embedding matrix* $E_d = [e_1, \ldots, e_{|\mathcal{B}|}] \in \mathbb{R}^{|\mathcal{B}| \times n}$. □

In Definition 5, E_d is randomly initialized, trainable, and shared in the economy-view sub-encoder.

Business-Structure Similarity: A business-structure similarity between a target district and an outbreak district is quantified by comparing their sales distributions with respect to *business categories*. Thus, the economy-view sub-encoder acquires a *business-structure representation* in Definition 6 by using the multi-head attention (Vaswani et al. 2017) to consider varying degrees of dependencies among business categories. For instance, the sales of "entertainment" are likely to depend on those of "leisure" but not on those of "logistics."

Definition 6. (BUSINESS-STRUCTURE REPRESENTATION) Let $X'_d = X_d - \Sigma_{i=1}^{|\mathcal{D}|} X_i / |\mathcal{D}|$ be the relative business structure feature of a target district d with a business structure feature X_d. Then, given the number h of attention heads, the dimensionality n of an embedding space, and a district-business embedding E_d, the *business-structure representation (BR)* is computed by

$$BR_{(d)} = [\text{Attn}_1(E_d)X'_d, \ldots, \text{Attn}_h(E_d)X'_d]^\top \in \mathbb{R}^{|\mathcal{B}| \times h},$$
$$\text{where} \quad \text{Attn}(E) = \text{Softmax}\big((EW_Q)(EW_K)^\top / \sqrt{n}\big), \quad (2)$$

W_Q and W_K are query and key projection matrices in $\mathbb{R}^{n \times (n/h)}$. □

Then, the *business-structure similarity* in Definition 7 is finally measured as the similarity between the business-structure representation of a target district and that of an outbreak district.

Definition 7. (BUSINESS-STRUCTURE SIMILARITY) Given a target district d and the outbreak district d_m of a mass infection case m, the *business-structure similarity* between d and d_m is calculated by

$$\mathrm{BS}(d, d_m) = \mathrm{Cosine}\big(BR_{(d)}, BR_{(d_m)}\big) \in \mathbb{R}^{|\mathcal{B}|}, \quad (3)$$

where $\mathrm{Cosine}(\cdot)$ is the cosine similarity function. □

Consumer-Structure Similarity: Contrary to the business categories, the consumer categories have less dependency on one another since they are all orthogonal perspectives (i.e., gender, age, and household). Thus, the economy-view sub-encoder directly calculates the *consumer-structure similarity* in Definition 8 based on the Jensen-Shannon divergence (JSD), which is one of the widely-used distribution divergence measures.

Definition 8. (CONSUMER-STRUCTURE SIMILARITY) Let X_d and X_{d_m} be the consumer structure features of a target district d and an outbreak district d_m, respectively. Then, the *consumer-structure similarity* between d and d_m is calculated by

$$\mathrm{CS}(d, d_m) = \mathrm{JSD}(X_d, X_{d_m}) \in \mathbb{R}^{|\mathcal{B}|}. \quad \square \quad (4)$$

Outbreak-Business Similarity: To determine the impact of the specific outbreak business of a mass infection case on businesses in the affected district, the economy-view sub-encoder calculates the *outbreak-business similarity* in Definition 9.

Definition 9. (OUTBREAK-BUSINESS SIMILARITY) Let E_d be the embedding matrix of a target district d and $e_m \in E_{d_m}$ be the embedding vector of the outbreak business in an outbreak district d_m. Then, the *outbreak-business similarity* between d and d_m is calculated by

$$\mathrm{OS}(d, d_m) = E_d(W_{os} \cdot e_m^\top) \in \mathbb{R}^{|\mathcal{B}|}, \quad (5)$$

where W_{os} is a trainable projection matrix in $\mathbb{R}^{n \times n}$. □

Finally, all the similarities are blended to generate an *economy-view representation* (ECR) for a target district d affected by an outbreak district d_m as the output of the economy-view sub-encoder,

$$\mathit{ECR}_{(d,d_m)} = \alpha \cdot \mathrm{BS}(d, d_m) + \beta \cdot \mathrm{CS}(d, d_m) + (1-\alpha-\beta) \cdot \mathrm{OS}(d, d_m), \quad (6)$$

where $\alpha \geq 0$ and $\beta \geq 0$ are the *trainable* weights to balance all the similarities such that $\alpha + \beta \leq 1$.

Geography-View Sub-Encoder An outbreak district has more economic impact on other districts if they are *geographically* or *socially* close. The geography-view sub-encoder encodes the geography-view feature into a *geography-view representation (GER)* containing these two types of closeness. For a target district d and an outbreak district d_m, GER is computed by

$$\mathit{GER}_{(d,d_m)} = \mathrm{FCN}\Big(\big[\mathrm{P\_Dist}(d, d_m), \mathrm{S\_Dist}(d, d_m)\big]\Big) \in \mathbb{R}, \quad (7)$$

where $\mathrm{FCN}(\cdot)$ is a fully connect neural network, $\mathrm{P\_Dist}(\cdot)$ and $\mathrm{S\_Dist}(\cdot)$ are the physical and social distance functions between two districts.

Epidemic-View Sub-Encoder The epidemic-view sub-encoder transforms the historical *epidemiological* severity feature into an upcoming economic severity. Specifically, we adopt the sequence encoder-decoder framework (Seq2Seq) (Bahdanau, Cho, and Bengio 2014). The LSTM encoder encodes the epidemic statistics sequence of an outbreak district d_m into the latent representation until the t-th day; the LSTM decoder decodes it to the *epidemic-view representation (EPR)* for the next w days,

$$\mathit{EPR}_{(d_m,w)} = \mathrm{FCN}\Big(\mathrm{Decoder}\big(\mathrm{Encoder}(d_m, t), t+w\big)\Big) \in \mathbb{R}^w, \quad (8)$$

where Encoder and Decoder are the LSTM encoder and decoder, respectively.

View Combiner The view combiner merges the repersentations from the three sub-encoders to produce the composite economic impact as a *microscopic representation (MIR)*,

$$\mathit{MIR}_{(d,d_m)} = \mathit{ECR}_{(d,d_m)} \otimes \mathit{GER}_{(d,d_m)} \otimes \mathit{EPR}_{(d_m,w)} \in \mathbb{R}^{|\mathcal{B}| \times w}, \quad (9)$$

where $\otimes$ is the outer product.

Phase 2: Macroscopic Aggregator

The macroscopic aggregator combines multiple microscopic representations of the target district, returned by the microscopic encoder for *all* outbreak districts. Thus, the *macroscopic representation (MAR)* is formulated by

$$\mathit{MAR}_{(d,\mathcal{M})} = \big[\mathit{MIR}_{(d,d_1)}, \ldots, \mathit{MIR}_{(d,d_{|\mathcal{M}|})}\big]^\top. \quad (10)$$

Last, the economic impact trend on all district-business pairs of the target district d, $\hat{\mathbb{Y}}_{(d,\,\cdot\,)}$, is finally predicted through a fully connected network with the gating mechanism that determines how much each outbreak affects the local economies in the district d, as formulated by

$$\begin{aligned} \hat{\mathbb{Y}}_{(d,\,\cdot\,)} &= \mathrm{FCN}\Big(\mathrm{Gate}\big(\mathit{MAR}_{(d,\mathcal{M})}\big)^\top \mathit{MAR}_{(d,\mathcal{M})}\Big) \\ &= \big[\hat{\mathbf{y}}_{(d,b_1)}, \hat{\mathbf{y}}_{(d,b_2)}, \ldots, \hat{\mathbf{y}}_{(d,b_{|\mathcal{B}|})}\big]^\top \in \mathbb{R}^{|\mathcal{B}| \times w}, \end{aligned} \quad (11)$$

where $\mathrm{Gate}(\cdot) = \mathrm{Softmax}(\mathrm{FCN}(\cdot))$ learns the aggregation weights to combine the economic impacts of the outbreaks on each target district-business pair.

Evaluation

Experiment Setup

Dataset Preparation The economy-view and geography-view features were derived from the economic activity dataset in 2019; the epidemic-view feature was derived from the mass infection dataset in 2020. The target variable, the economic impact trend in 2020, was derived from the economic activity dataset. Because the mass infection cases were collected from Seoul, all 25 districts in Seoul were included in the experiments. We divided the epidemic-view feature and the economic impact trend into two periods: from February 2020 to October 2020 for the training set and November 2020 for the test set. For the prediction time span, we used 14 days ($w = 14$) and 28 days ($w = 28$) for short-term and mid-term forecasting, respectively. That is, the period of November 1–14 and that of November 1–28 were predicted at the point of the end of the training set (October 31, 2020).

Method	District	Business Category: Leisure	Fashion Goods	Travel	Logistics	Restaurant	Cram School	Pubs&Bars	Cultural Activity
Seq2Seq +Attn	Jung	0.44 (±0.03)	1.13 (±0.04)	0.69 (±0.03)	1.22 (±0.03)	1.38 (±0.03)	0.47 (±0.09)	1.10 (±0.07)	1.05 (±0.03)
	Jungnang	0.38 (±0.03)	0.47 (±0.11)	0.21 (±0.02)	0.31 (±0.03)	0.30 (±0.03)	0.49 (±0.01)	0.37 (±0.04)	0.29 (±0.02)
	Mapo	0.25 (±0.05)	0.37 (±0.06)	0.85 (±0.03)	0.72 (±0.07)	**0.56** (±0.03)	0.54 (±0.05)	0.62 (±0.04)	1.00 (±0.03)
TCN	Jung	0.39 (±0.03)	1.07 (±0.03)	0.62 (±0.04)	1.15 (±0.04)	1.32 (±0.04)	0.24 (±0.02)	1.14 (±0.08)	0.99 (±0.04)
	Jungnang	0.50 (±0.06)	0.44 (±0.07)	0.19 (±0.01)	0.25 (±0.02)	0.27 (±0.04)	0.45 (±0.01)	0.40 (±0.04)	0.34 (±0.04)
	Mapo	0.28 (±0.08)	0.33 (±0.03)	0.78 (±0.04)	0.65 (±0.06)	0.61 (±0.02)	0.37 (±0.07)	0.67 (±0.02)	0.94 (±0.04)
TADA	Jung	0.30 (±0.05)	1.06 (±0.02)	0.37 (±0.06)	0.98 (±0.15)	0.49 (±0.05)	0.40 (±0.06)	1.32 (±0.06)	0.39 (±0.08)
	Jungnang	0.46 (±0.06)	0.75 (±0.06)	0.24 (±0.05)	0.36 (±0.05)	0.25 (±0.03)	0.48 (±0.04)	0.41 (±0.03)	0.33 (±0.03)
	Mapo	0.41 (±0.06)	0.41 (±0.06)	0.39 (±0.10)	0.54 (±0.10)	0.61 (±0.07)	0.62 (±0.04)	0.80 (±0.06)	0.45 (±0.10)
DEFSI	Jung	**0.23** (±0.03)	1.04 (±0.01)	0.34 (±0.04)	0.64 (±0.02)	0.38 (±0.08)	**0.22** (±0.03)	1.15 (±0.05)	0.32 (±0.06)
	Jungnang	0.51 (±0.02)	0.51 (±0.03)	**0.17** (±0.01)	**0.25** (±0.00)	0.21 (±0.02)	0.36 (±0.03)	0.42 (±0.03)	0.27 (±0.02)
	Mapo	0.28 (±0.04)	0.52 (±0.04)	0.29 (±0.01)	0.63 (±0.03)	0.61 (±0.05)	0.32 (±0.04)	0.66 (±0.04)	0.55 (±0.06)
COVID-EENet	Jung	0.30 (±0.04)	**1.01** (±0.01)	**0.16** (±0.03)	**0.32** (±0.04)	**0.14** (±0.03)	**0.22** (±0.03)	**0.78** (±0.04)	**0.22** (±0.03)
	Jungnang	**0.33** (±0.03)	**0.35** (±0.03)	0.41 (±0.03)	0.26 (±0.03)	**0.17** (±0.01)	**0.23** (±0.03)	**0.25** (±0.02)	**0.25** (±0.02)
	Mapo	**0.20** (±0.04)	**0.28** (±0.02)	**0.19** (±0.04)	**0.32** (±0.05)	**0.56** (±0.03)	**0.26** (±0.03)	**0.34** (±0.06)	**0.27** (±0.03)

Table 3: RMSE for 14 days prediction ($w = 14$) in representative district-business pairs.

Algorithms and Evaluation Metrics For comparison with *COVID-EENet*, we chose four deep learning-based algorithms for sequence modeling that can be employed in our problem setting—two time-series prediction models *Seq2Seq+Attn* (Bahdanau, Cho, and Bengio 2014) and *temporal convolutional network (TCN)* (Bai, Kolter, and Koltun 2018), a sales prediction model *TADA* (Chen et al. 2018), and an epidemiological model *DEFSI* (Wang, Chen, and Marathe 2019). See Section C of the supplementary material for the details of the baselines and implementations.

For fair comparison, the exactly *same* features were fed to *COVID-EENet* and the four baselines. The baselines were modified to use the same features: the economy-view feature and the geography-view feature were concatenated and then encoded using fully-connected networks; the impact of concurrent outbreaks was obtained by simply averaging the impact of an individual outbreak because there is no module corresponding to the macroscopic aggregator. For reproducibility, the source code of *COVID-EENet* and the baselines is available at https://bit.ly/covideenet.

We used the *root mean squared error (RMSE)* and the *mean absolute error (MAE)* to evaluate the algorithms. The mean and standard error of *five* repetitions with different random initialization were reported.

Overall Performance Comparison

Table 3 presents the RMSE results for upcoming 14 days in representative district-business pairs. Three districts, "Jung," "Jungnang," and "Mapo," were chosen from each of dominant functional district types—commercial & business, residential, and diversified, classified by the Ministry of Environment, Korea. Then, seven business categories with high sales amount were selected for the overall comparison. See Section D of the supplementary material for other district-business pairs. Overall, *COVID-EENet* achieved the highest accuracy (lowest RMSE) for most district-business pairs and thus showed the versatility regardless of district functional types and business categories.

Table 4 presents the RMSE and MAE results for upcoming 14 and 28 days, averaged over all $850 (= 25 \times 34)$ district-business pairs. As expected, *COVID-EENet* outperformed the four existing algorithms by a large margin: 6.5–9.3% in terms of RMSE and 9.3–15.1% in terms of MAE.

Method	14 days ($w = 14$) RMSE	14 days ($w = 14$) MAE	28 days ($w = 28$) RMSE	28 days ($w = 28$) MAE
Seq2Seq+Attn	0.442 (±0.003)	0.258 (±0.003)	0.471 (±0.004)	0.285 (±0.003)
TCN	0.444 (±0.004)	0.261 (±0.002)	0.473 (±0.003)	0.284 (±0.002)
TADA	0.448 (±0.006)	0.262 (±0.006)	0.468 (±0.003)	0.277 (±0.002)
DEFSI	0.428 (±0.003)	0.246 (±0.004)	0.450 (±0.001)	0.264 (±0.001)
COVID-EENet	**0.403** (±0.002)	**0.223** (±0.003)	**0.437** (±0.003)	**0.254** (±0.003)
AVG Improv.	9.3%	15.1%	6.5%	9.3%

Table 4: RMSE and MAE for 14 and 28 days, averaged over "all" district-business pairs.

The improvements over the existing algorithms are sufficiently high for *both* forecast horizons; the improvement for 14-day prediction is more noticeable than that for 28-day prediction since a longer forecast horizon is more likely affected by external factors unknown at the moment.

Achieving high accuracy for the fine-grained EEM requires (i) finding the complex relationships from the *multi-view*—economy, geography, and epidemic—features to the economic impact trend and (ii) handling *concurrent* mass infection cases altogether. In *COVID-EENet*, the two-level architecture can support these two requirements. Specifically, the three sub-encoders in the microscopic encoder fulfill the former requirement, and the macroscopic aggregator satisfies the latter requirement. On the contrary, the existing algorithms are not designed to support the two requirements, though we chose them to be closest to our problem; therefore, multi-view features and concurrent outbreaks cannot be fully exploited in the existing algorithms.

Ablation Study

We conducted ablation studies using the three variants in Table 5, each of which lacks one of the main components—the *economy-view sub-encoder* and the *geography-view sub-encoder* in the microscopic encoder and the *macroscopic aggregator* itself. The epidemic-view sub-encoder is essential for the problem setting and thus could not be excluded. Table 5 summarizes the results for the three variants.

Effect of Microscopic Encoder *COVID-EENet* (w/o economy-view) faced the sharpest drop in the accuracy, empirically showing that considering the similarity in terms of business categories is indeed fundamental to modeling the impact of mass infection cases. Meanwhile, the result

Variants	RMSE	Degrade	MAE	Degrade
No Economy-View	0.502 (±0.005)	12.9%	0.297 (±0.002)	14.5%
No Geography-View	0.443 (±0.002)	1.4%	0.258 (±0.003)	1.6%
No Macroscopic Agg	0.444 (±0.006)	1.6%	0.259 (±0.004)	1.9%
COVID-EENet	0.437 (±0.003)	–	0.254 (±0.003)	–

Table 5: Ablation study on main components ($w = 28$).

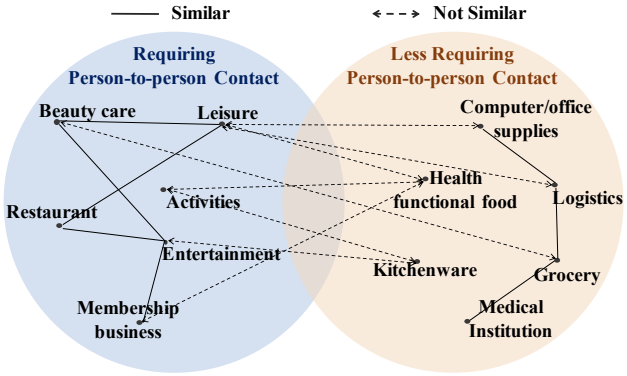

Figure 3: Business disparities captured by the ECR.

of *COVID-EENet* (w/o geography-view) shows that the geography view was less important than the economy view, possibly because the road and public transport infrastructure in Seoul is well-equipped.

Effect of Macroscopic Aggregator *COVID-EENet* (w/o macroscopic aggregator) is, in fact, identical to simply averaging the economic impacts from all mass infections as in the baselines, thus not providing an optimal aggregation. The performance degradation means that the macroscopic aggregator is required to precisely determine the contribution of each mass infection case. The performance of this variant is still higher than those of the baselines, because it is equipped with the complete *microscopic encoder* which better reflects the *multi-view* aspects.

Case Study and Discussion

We discuss *business disparities* among the three insights summarized in Introduction. The discussion about district-business and epidemic disparities is included in Section E of the supplementary material.

Business Disparities: The determinant contributing to business disparities is how much *person-to-person contact* is required. People try to avoid closed spaces, crowded places, and close-contact as much as possible. We note that *COVID-EENet* captures this determinant by averaging *district-business embedding matrix* E_d in Eq. (5) along districts. Figure 3 shows the clustering results for business categories by spectral clustering (Ng, Jordan, and Weiss 2002), where a solid or dashed edge indicates being close or far according to the Euclidean distance in the embedding space. Interestingly, a group (left) mostly contains the businesses that *significantly* require person-to-person contact, while the other group (right) mostly contains those that do not. As confirmed in Section A (supplementary material), the business categories in the left group mostly suffer from sales decline, but those in the right group even benefit from the pandemic.

Related Work

COVID-19 has gained a lot of attention from researchers to alleviate its devastating impact (Shorten, Khoshgoftaar, and Furht 2021; Nguyen 2020; Hussain et al. 2020), especially by predicting the epidemic trends and economic impacts.

Epidemic Models

Numerous studies exploited a traditional approach called the susceptible exposed infected resistant (SEIR) model to predict the spread of COVID-19 (Dandekar and Barbastathis 2020; Arik et al. 2020; He, Peng, and Sun 2020; Annas et al. 2020; Pandey et al. 2020; Chang et al. 2021). Also, many studies enjoyed the power of the DNN to predict the spread of COVID-19 and its impact. Zeroual et al. (2020) suggested a model with five different RNNs and a VAE to predict the spread of COVID-19. Kim et al. (2020) predicted the number of inbound COVID-19 patients via an architecture based on the geographic hierarchy. Pal et al. (2020) designed a country-specific network and predicted risk using the LSTM model and the Bayesian optimization framework. Despite these previous studies predicting the spread of COVID-19, we note that its economic impact is *not* considered yet.

Economic Models

A number of studies have been developed for time-series prediction related to the economy (Chen et al. 2018; Seeger, Salinas, and Flunkert 2016; Kim 2003). In particular, Chen et al. (2018) proposed a novel framework, named TADA, using trend alignment-based multitask RNNs with dual-attention; the framework improved the performance of sales prediction tasks by aligning the upcoming trend with relevant historical trends. Some studies have developed the methods of analyzing the economic impact caused by infectious diseases through traditional statistical modeling (Ahmar and del Val 2020; Meltzer, Cox, and Fukuda 1999; Lee et al. 2013; Berger, Herkenhoff, and Mongey 2020). In particular, Berger, Herkenhoff, and Mongey (2020) extended the SEIR COVID-19 model to understand the role of testing and quarantine; they showed that testing at a higher rate with targeted quarantine policies can dampen the economic impact of COVID-19. As far as we know, there has been no such study predicting the fine-grained economic impact of infectious diseases.

Conclusion

In this paper, we have proposed a novel approach to predicting the *fine-grained* impact of COVID-19 on local economies. We are provided with an aggregated credit card transaction dataset by the courtesy of the BC Card corporation and carefully derive the three fine-grained features which represent the key factors of economic-epidemiological modeling. ***COVID-EENet*** incorporates these features through its two-level architecture and can predict the economic impacts caused by concurrent mass infection cases very accurately. This work is the first work to bridge the gap between mass infection cases and economic activities in local businesses. Overall, we expect that the government authorities will be able to construct proactive financial aids for the fragile public with the help of the fine-grained prediction result.

Acknowledgement

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00862, DB4DL: High-Usability and Performance In-Memory Distributed DBMS for Deep Learning).

References

Acs, G.; and Karpman, M. 2020. Employment, Income, and Unemployment Insurance during the COVID-19 Pandemic. Technical report, Urban Institute.

Ahmar, A. S.; and del Val, E. B. 2020. SutteARIMA: Short-term Forecasting Method, a Case: COVID-19 and Stock Market in Spain. *Science of the Total Environment*, 729: 138883.

Annas, S.; Pratama, M. I.; Rifandi, M.; Sanusi, W.; and Side, S. 2020. Stability Analysis and Numerical Simulation of SEIR Model for Pandemic COVID-19 Spread in Indonesia. *Chaos, Solitons & Fractals*, 139: 110072.

Arik, S. O.; Li, C.-L.; Yoon, J.; Sinha, R.; Epshteyn, A.; Le, L. T.; Menon, V.; Singh, S.; Zhang, L.; Yoder, N.; et al. 2020. Interpretable Sequence Learning for COVID-19 Forecasting. ArXiv:2008.00646.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. In *ICLR*.

Bai, S.; Kolter, J. Z.; and Koltun, V. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. ArXiv:1803.01271.

Berger, D. W.; Herkenhoff, K. F.; and Mongey, S. 2020. An SEIR Infectious Fisease Model with Testing and Conditional Quarantine. Technical report, National Bureau of Economic Research.

Callinan, S.; and MacLean, S. 2020. COVID-19 Makes a Stronger Research Focus on Home Drinking More Important than Ever. *Drug and Alcohol Review*, 39(6): 613–615.

Chang, S.; Pierson, E.; Koh, P. W.; Gerardin, J.; Redbird, B.; Grusky, D.; and Leskovec, J. 2021. Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening. *Nature*, 589: 82–87.

Chen, T.; Yin, H.; Chen, H.; Wu, L.; Wang, H.; Zhou, X.; and Li, X. 2018. TADA: Trend Alignment with Dual-Attention Multi-Task Recurrent Neural Networks for Sales Prediction. In *ICDM*, 49–58.

Cheng, C.; Barceló, J.; Hartnett, A. S.; Kubinec, R.; and Messerschmidt, L. 2020. COVID-19 Government Response Event Dataset (CoronaNet v. 1.0). *Nature Human Behaviour*, 4(7): 756–768.

Coibion, O.; Gorodnichenko, Y.; and Weber, M. 2020. Labor Markets During the COVID-19 Crisis: A Preliminary View. Technical report, National Bureau of Economic Research.

Dandekar, R.; and Barbastathis, G. 2020. Quantifying the Effect of Quarantine Control in COVID-19 Infectious Spread Using Machine Learning. MedRxiv.

Demirguc-Kunt, A.; Lokshin, M.; and Torre, I. 2020. Opening-up Trajectories and Economic Recovery: Lessons after the First Wave of the COVID-19 Pandemic. Technical report, World Bank Group.

Di Clemente, R.; Luengo-Oroz, M.; Travizano, M.; Xu, S.; Vaitla, B.; and González, M. C. 2018. Sequences of Purchases in Credit Card Data Reveal Lifestyles in Urban Populations. *Nature Communications*, 9(3330): 1–8.

Elgin, C.; Basbug, G.; and Yalaman, A. 2020. Economic Policy Responses to a Pandemic: Developing the COVID-19 Economic Stimulus Index. *Covid Economics*, 1(3): 40–53.

Hale, T.; Petherick, A.; Phillips, T.; and Webster, S. 2020. Variation in Government Responses to COVID-19. Technical report, Blavatnik School of Government.

Harris, J. E. 2020. Geospatial Analysis of the September 2020 Coronavirus Outbreak at the University of Wisconsin-Madison: Did a Cluster of Local Bars Play a Critical Role? Technical report, National Bureau of Economic Research.

He, S.; Peng, Y.; and Sun, K. 2020. SEIR Modeling of the COVID-19 and Its Dynamics. *Nonlinear Dynamics*, 101(3): 1667–1680.

Hussain, A. A.; Bouachir, O.; Al-Turjman, F.; and Aloqaily, M. 2020. AI Techniques for COVID-19. *IEEE Access*, 8: 128776–128795.

Kim, K.-j. 2003. Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing*, 55(1-2): 307–319.

Kim, M.; Kang, J.; Kim, D.; Song, H.; Min, H.; Nam, Y.; Park, D.; and Lee, J.-G. 2020. Hi-COVIDNet: Deep Learning Approach to Predict Inbound COVID-19 Patients and Case Study in South Korea. In *KDD*, 3466–3473.

Lee, B. Y.; Bacon, K. M.; Bottazzi, M. E.; and Hotez, P. J. 2013. Global Economic Burden of Chagas Disease: A Computational Simulation Model. *The Lancet Infectious Diseases*, 13(4): 342–348.

Meltzer, M. I.; Cox, N. J.; and Fukuda, K. 1999. The Economic Impact of Pandemic Influenza in the United States: Priorities for Intervention. *Emerging Infectious Diseases*, 5(5): 659.

Neumann-Böhme, S.; Varghese, N. E.; Sabat, I.; Barros, P. P.; Brouwer, W.; van Exel, J.; Schreyögg, J.; and Stargardt, T. 2020. Once We Have It, Will We Use It? A European Survey on Willingness To Be Vaccinated against COVID-19. *The European Journal of Health Economics*, 21: 977–982.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On Spectral Clustering: Analysis and an Algorithm. In *NIPS*, 849–856.

Nguyen, T. T. 2020. Artificial Intelligence in the Battle Against Coronavirus (COVID-19): A Survey and Future Research Directions. ArXiv:2008.07343.

Pal, R.; Sekh, A. A.; Kar, S.; and Prasad, D. K. 2020. Neural Network Based Country Wise Risk Prediction of COVID-19. *Applied Sciences*, 10(18): 6448.

Pandey, G.; Chaudhary, P.; Gupta, R.; and Pal, S. 2020. SEIR and Regression Model based COVID-19 Outbreak Predictions in India. ArXiv:2004.00958.

Rahman, M. A.; Zaman, N.; Asyhari, A. T.; Al-Turjman, F.; Bhuiyan, M. Z. A.; and Zolkipli, M. 2020. Data-Driven Dynamic Clustering Framework for Mitigating the Adverse Economic Impact of COVID-19 Lockdown Practices. *Sustainable Cities and Society*, 62: 102372.

Seeger, M.; Salinas, D.; and Flunkert, V. 2016. Bayesian Intermittent Demand Forecasting for Large Inventories. In *NIPS*, 4653–4661.

Shorten, C.; Khoshgoftaar, T. M.; and Furht, B. 2021. Deep Learning Applications for COVID-19. *Journal of Big Data*, 8(1): 1–54.

Sidhu, G. S.; Rai, J. S.; Khaira, K. S.; Kaur, S.; et al. 2020. The Impact of COVID-19 Pandemic on Different Sectors of the Indian Economy: A Descriptive Study. *International Journal of Economics and Financial Issues*, 10(5): 113–120.

Singh, A. 2020. COVID-19 Pandemic and the Future of SDGs. In Malhotra, V. K.; Fernando, R. L. S.; and Haran, N. P., eds., *Disaster Management for 2030 Agenda of the SDG*, 279–317.

Sumner, A.; Hoy, C.; Ortiz-Juarez, E.; et al. 2020. Estimates of the Impact of COVID-19 on Global Poverty. Technical report, United Nations University World Institute for Development Economics Research.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. In *NIPS*, 5998–6008.

Wang, L.; Chen, J.; and Marathe, M. 2019. DEFSI: Deep Learning Based Epidemic Forecasting with Synthetic Information. In *AAAI*, 9607–9612.

Zeroual, A.; Harrou, F.; Dairi, A.; and Sun, Y. 2020. Deep Learning Methods for Forecasting COVID-19 Time-Series Data: A Comparative Study. *Chaos, Solitons & Fractals*, 140: 110121.