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

Do Young Kim\(^1\), Hyang Suk Min\(^1\), Youn Geun Nam\(^1\), Hwan Jun Song\(^2\), Susik Yoon\(^3\), Min Seok Kim\(^1\), Jaegil Lee\(^1\)*

\(^1\)KAIST, \(^2\)NAVER AI Lab, \(^3\)University of Illinois at Urbana-Champaign
\{dodokim, hyangsk.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://github.com/kaist-dmlab/COVID-EENet.

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. Instead of deliberating 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 Yalamov 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 enables 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. 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.

### Economic-Epidemiological Data

#### Dataset Description

**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. 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 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) ×

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Transaction date (YYYYMMDD)</td>
</tr>
<tr>
<td>Address</td>
<td>Store district</td>
</tr>
<tr>
<td>Business kind</td>
<td>Store business</td>
</tr>
<tr>
<td>Nationality</td>
<td>Korean or foreigner</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, female, or others</td>
</tr>
<tr>
<td>Age</td>
<td>Age groups in 10s</td>
</tr>
<tr>
<td>Household type</td>
<td>Classification by family size, etc.</td>
</tr>
<tr>
<td>Price</td>
<td>Total price of the sales</td>
</tr>
<tr>
<td>Count</td>
<td>Total count of the sales</td>
</tr>
</tbody>
</table>

Table 1: Key attributes in economic activity data.
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. 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). Each mass infection case typically resulted in over 100 confirmed cases.

**Exploratory Data Analysis**

In order to ascertain the underlying claims of this study, we conducted an exploratory data analysis using the economic-epidemiological dataset.

**Economy Decline and Degree of Impact** First of all, we want to answer the question: “Does the COVID-19 pandemic damage the economy?” Figure 2(a) shows the changes in the sales for several business categories in 2020 (after the pandemic) compared with 2019 (before the pandemic). Overall, the sales declined by 2.4% across all districts and business categories in Seoul. However, the impact significantly differed by business categories. In general, the businesses vulnerable to the virus infection, e.g., “restaurant” and “entertainment,” were severely damaged, whereas “grocery” and “logistics” even benefited from the pandemic.

Then, let us answer the question: “Is the economic impact diverse depending on the business category and district?” Figure 2(b) shows the monthly sales of two business categories, each of which was gained or damaged by the pandemic. The monthly sales of both categories were rather stable in 2019; however, those of “restaurant” in 2020 showed a sharp decline in February, whereas those of “grocery” in 2020 generally increased throughout the year. Thus, this example shows the diversity of the economic impact across business categories. Figure 2(c) shows the daily sales in two districts for the “entertainment” business in five weeks from March 2, 2020. The sales in the “Gangnam” district fluctuated a lot, whereas those in the “Guro” district did not. Thus, this example shows another diversity across districts.

**Need for Fine Granularity** “Is the fine-grained economic-epidemiological modeling (EEM) indeed necessary?” The diverse degree of the impact clearly gives an answer to this question. In Figure 2(d), the daily sales averaged over all districts are also presented in blue. The average trend obviously does not reflect the individual trends in the two districts.

**Impact of Mass Infection** Last, we answer the important question: “Does a mass infection case affect the economic activity in the vicinity?” Figure 2(d) shows the daily sales in the “Guro” district for the “entertainment” business in March and April 2020, with the dates when two mass infection cases occurred in that district. We compared the same day of the week, Friday, which typically reached its peak in that business. The sales dropped immediately after the two mass infection cases and started to recover in a couple of weeks. Since mass infection cases are reported frequently in public media, they significantly affect consumer sentiment.

**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):

\[
\langle \ldots, \text{fraction of the sales for the } i\text{-th business category}, \ldots \rangle.
\]

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)\(^4\):

\[
\langle \ldots, \text{fraction of the sales by the } j\text{-th consumer category}, \ldots \rangle.
\]

Again, each fraction is the average of the two fractions created in terms of price and count.

\(^4\)3 (gender) \times 3 (age) \times 3 (household) = 27.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin place</td>
<td>District the mass infection occurred</td>
</tr>
<tr>
<td>Business category</td>
<td>Business category the mass infection occurred</td>
</tr>
<tr>
<td>Title</td>
<td>Known title of the mass infection</td>
</tr>
<tr>
<td>Start date</td>
<td>Date the mass infection started</td>
</tr>
<tr>
<td>End date</td>
<td>Date the mass infection ended</td>
</tr>
<tr>
<td>Confirmed cases</td>
<td>Daily number of confirmed cases</td>
</tr>
</tbody>
</table>

Table 2: Key attributes in the mass infection data.

---

\(^3\)https://www.seoul.go.kr/

\(^4\)http://www.kdca.go.kr/
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 invariant 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.

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, \( \{ \text{number of confirmed cases} (i) \text{ in a specific day}, \text{ (ii) within a week}, \text{ (iii) until the specific day}, \text{ (iv) number of elapsed days since the beginning of the mass infection} \} \ldots \}. \) The length of the sequence is determined by the duration of the corresponding mass infection case.

Given a set \( D \) of districts and a set \( B \) 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 \( (d, b) \in D \times B \), where \( d \in D \) and \( b \in B \).

A set \( M \) of mass infection cases is given as the source of the economic impact, where each mass infection case \( m \in M \) is specified by Definition 2.

**Definition 2. (Mass Infection Case)** A mass infection case \( m \in M \) 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 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 \( y_{(d,b)} = \{y(t)\}_{t=1}^w \), where

\[
y(t) = \frac{\text{sales amount on day } t - \text{sales amount a year ago from day } t}{\text{sales amount a year ago from day } t}
\]

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 \( D \) of districts, a set \( B \) of business categories, and a set \( M \) of mass infection cases, to predict the economic impact trend \( y_{(d,b)} \) in Definition 3 for each district-business pair \( (d,b) \in D \times B \) 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 3 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 \( EPR \), respectively.
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.

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 \(e_{db} \in \mathbb{R}^n\) 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 \(E_d = [e_1, \ldots, e_{|B|}] \in \mathbb{R}^{|B| \times n}\).

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 \(X'_d = X_d - \sum_{n=1}^{|B|} X_n/|D|\) 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

\[
BR_d = [\text{Attn}_1(E_d)X'_d, \ldots, \text{Attn}_n(E_d)X'_d] \in \mathbb{R}^{[8] \times h},
\]

where \(\text{Attn}(E) = \text{Softmax}((EW_Q)(EW_K)^\top/\sqrt{n})\).

\(W_Q\) 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

\[
BS(d, d_m) = \text{Cosine}(BR_d, BR_{d_m}) \in \mathbb{R}^{[8]},
\]

where \(\text{Cosine}(\cdot)\) 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

\[
CS(d, d_m) = \text{JSD}(X_d, X_{d_m}) \in \mathbb{R}^{[8]}.
\]

Outbreak-Business Similarity: To determine the impact of the specific outbreak business of a mass infection case on local 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

\[
OS(d, d_m) = E_d(W_{os} \cdot e_m^\top) \in \mathbb{R}^{[8]},
\]

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,

\[
ECR_{(d, d_m)} = \alpha \cdot BS(d, d_m) + \beta \cdot CS(d, d_m) + \gamma \cdot OS(d, d_m),
\]

where \(\alpha, \beta, \gamma \geq 0\) are the trainable weights to balance all the similarities such that \(\alpha + \beta + \gamma = 1\).
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

$$\text{GER}_{(d,d_m)} = \text{FCN} \left( \left[ \text{P} \cdot \text{Dist}(d,d_m), \text{S} \cdot \text{Dist}(d,d_m) \right] \right) \in \mathbb{R}^d,$$

(7)

where FCN(·) is a fully connect neural network, P·Dist(·) and S·Dist(·) 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,

$$\text{EPR}_{(d_m,w)} = \text{FCN} \left( \text{Decoder} (\text{Encoder}(d_m,t), t+w) \right) \in \mathbb{R}^w,$$

(8)

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

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

$$\text{MIR}_{(d,d_m)} = \text{ECR}_{(d,d_m)} \otimes \text{GER}_{(d,d_m)} \otimes \text{EPR}_{(d_m,w)} \in \mathbb{R}^{(|M|) \times w},$$

(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

$$\text{MAR}_{(d,M)} = \text{MIR}_{(d,d_1)} \ldots \text{MIR}_{(d,d_{|M|})}^T.$$

(10)

Last, the economic impact trend on all district-business pairs of the target district $d$, $\hat{\mathbf{Y}}(d, \cdot)$, is predicted through a FCN with the gating mechanism determining the contributions of outbreaks to the local economies, as formulated by

$$\hat{\mathbf{Y}}(d, \cdot) = \text{FCN} \left( \text{Gate} \left( \text{MAR}_{(d,M)} \right)^T \text{MAR}_{(d,M)} \right)$$

$$= \left[ \hat{Y}(d,b_1), \hat{Y}(d,b_2), \ldots, \hat{Y}(d,b_{|B|}) \right]^T \in \mathbb{R}^{(|B|) \times w},$$

(11)

where Gate(·) = Softmax(FCN(·)) learns the aggregation weights to combine the economic impacts of the outbreaks on each target district-business pair.

Training Algorithm Pseudocode

The training procedure of COVID-EENet is described in Algorithm 1. The training algorithm receives the district-business pairs, the mass infection cases, the three view features, and the target economic impact trend; and returns the optimal parameter. The forward propagation is carried out through the microscopic encoder (Lines 7–13) and the macroscopic aggregator (Lines 14–16). In the microscopic encoder, the three sub-encoders are concurrently executed (Lines 9–11), and the learned representations for each view are combined to form the macroscopic representation (Line 12). In the macroscopic aggregator, after all microscopic representations are concatenated to form the macroscopic representation, the gating module is applied to predict the economic impact trend as the target variable (Line 16). Last, based on the RMSE loss, the backward propagation is carried out to update the network parameter (Lines 18–19).

Algorithm 1: COVID-EENet Training

INPUT: District $D$, business $B$, mass infection $M$
economy-view $X_\nu$, geography-view $X_G$, epidemic-view $X_\nu$, economic impact trend $Y$

OUTPUT: Optimal model parameters $\Theta^*$

1: $\Theta$ ← Initialize model parameters;
2: for epoch = 1 to epochs do
3:   for $t = 1$ to $\#$ of training days do
4:     $M_t$ ← subset of $M$ effective at day $t$;
5:     for each $d \in D$ do
6:       $k ← 1$; /* mass infection index */
7:       /* PHASE 1: MICROSCOPIC ENCODER */
8:       for each $m ∈ M_t$ do
9:         $\text{ECR}_k ← \text{EconomySubEncoder}(d,m,X_\nu)$;
10:        $\text{GER}_k ← \text{GeographySubEncoder}(d,m,X_G)$;
11:       $\text{EPR}_k ← \text{EpidemicSubEncoder}(d,m,X_\nu)$;
12:       $\text{MIR}_k ← \text{ViewCombiner}(\text{ECR}_k, \text{GER}_k, \text{EPR}_k)$;
13:     $k ← k + 1$;
14:     /* PHASE 2: MACROSCOPIC AGGREGATOR */
15:     $\text{MAR} ← (\text{MIR}_1, \ldots, \text{MIR}_{|M_t|})$;
16:     $\hat{\mathbf{Y}}(d, \cdot) ← \text{FCN-Gate}(\text{MAR})$;
17: /* MODEL UPDATE */
18:     Compute the loss $\mathcal{J}$ using $\mathbf{Y}(d, \cdot)$ and $\hat{\mathbf{Y}}(d, \cdot)$;
19:     $\Theta^* ← \Theta^* − α \nabla \mathcal{J}$;
20: return $\Theta^*$;

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 periods of November 1–14 and November 1–28 were predicted at the point of the end of the training set (October 31, 2020).
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, Koltur, and Koltun 2018), a sales prediction model TADA (Chen et al. 2018), and an epidemiological model DEFSI (Wang, Chen, and Marathe 2019).

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. The source code of COVID-EENet and the baselines is available at https://github.com/kaist-dmlab/COVID-EENet.

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, six business categories with high sales amount were selected for the overall comparison. 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” district-business pairs. 9.3% in terms of RMSE and 9.3–15.1% in terms of MAE. 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.
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 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.

Detailed Analysis on Economic-View  

We analyze the components of the economic-view sub-encoder considering its significant contribution to the performance. Regarding the contribution of ECR’s sub-components, inspection of the trainable weights in Eq. (6) indicates that the contribution of business-structure similarity (BS) is higher than those of consumer-structure (CS) and outbreak-business similarity (OS); specifically, the contributions of BS, CS, and OS are 36%, 32.5%, and 31.5%, respectively. Regarding the effect of the attention in business-structure representation, because the attention in Eq. (2) is required to capture varying degrees of the dependencies among businesses, its removal naturally degrades the accuracy by 1.9–3.6% as shown in Table 6.

Case Study and Discussion  

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 Def. (5) along districts. Figure 4 shows the clustering results for business categories by spectral clustering (Ng,

<table>
<thead>
<tr>
<th>Variants</th>
<th>RMSE</th>
<th>Degrade</th>
<th>MAE</th>
<th>Degrade</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Econ-View</td>
<td>0.502</td>
<td>12.9%</td>
<td>0.297</td>
<td>14.5%</td>
</tr>
<tr>
<td>No Geog-View</td>
<td>0.443</td>
<td>1.4%</td>
<td>0.258</td>
<td>1.6%</td>
</tr>
<tr>
<td>No Macro Agg</td>
<td>0.444</td>
<td>1.6%</td>
<td>0.259</td>
<td>1.9%</td>
</tr>
<tr>
<td><em>COVID-EENet</em></td>
<td>0.437</td>
<td>–</td>
<td>0.254</td>
<td>–</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Variants</th>
<th>RMSE</th>
<th>Degrade</th>
<th>MAE</th>
<th>Degrade</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Attn (Eq. (2))</td>
<td>.441</td>
<td>1.9%</td>
<td>.231</td>
<td>3.6%</td>
</tr>
<tr>
<td><em>COVID-EENet</em></td>
<td>.403</td>
<td>–</td>
<td>.223</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 6: Ablation study on the attention mechanism of business-structure representations ($w = 14$).

*Figure 4: Business disparities captured by the ECR.*

*Figure 5: District-business disparities in grocery businesses.*

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 exploratory data analysis, the business categories in the left group mostly suffer from sales decline, but those in the right group even benefit from the pandemic.

**District-Business Disparities**  

The fine-grained determinant contributing to district-business disparities is whether a district’s functionality is residential or commercial. People reduce unnecessary movements to minimize the risk of infection; thus, economic activities could be confined to their residential areas. To ascertain this insight, we choose the “grocery” category, in which the sales amount generally increased in 2020 compared with that in 2019, as an example. Figure 5(a) shows the sales changes of two districts in August, 2020 when mass infection cases concurrently occurred. The sales amount in a residential district (e.g., “Dobong”) generally increased as shown by positive sales changes, but the sales amount in a commercial district (e.g., “Dongdaemun”) generally decreased as shown by negative sales changes. *COVID-EENet* can capture this determinant by deriving the economy-view representation (ECR). In Figure 5(b), we plot the value of the “grocery” dimension in the ECR for several districts. Interestingly, the ECR values were mostly negative for commercial districts whose local grocery businesses were severely damaged, whereas those values were mostly positive for residential districts.
COVID-19 has gained a lot of attention to alleviate its devastating impact (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; 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. However, there has been no such study predicting the fine-grained economic impact of infectious diseases.

Related Work

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.
Acknowledgements

This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2020R1A2B5B03095947) and 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


