Analysis of the Temporal Structure in Economic Condition Assessments

Masahiro Kato
Data Analytics Department
Mizuho-DL Financial Technology Co., Ltd.
Tokyo, Japan
masahiro-kato@fintec.co.jp

2024 16th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI) | 979-8-3503-7790-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IIAI-AAI63651.2024.00065

***Abstract*—The *Economy Watcher Survey*, a market survey published by the Japanese government, features *assessments of current and future economic conditions* made by individuals from various occupations. While this survey offers crucial insights for economic policy formulation, the specific timing of the events that form the basis for these assessments remains unclear. For instance, one individual might assess future economic conditions based on events that have already occurred, whereas another might base their judgment on potential future events. Understanding the temporal context in which people base their assessments is vital for policymakers. Motivated by this need, we aim to clarify the *different points in time that underpin assessments of future economic conditions*. In our research, we categorize assessments of future economic conditions into *assessments based on current and future events*, employing *learning from positive and unlabeled data* (*PU learning*). Given that the dataset spans multiple periods, we introduce a novel architecture that enables neural networks to perform PU learning efficiently, drawing on principles of *multi-task learning*. Our empirical analysis validates that our method effectively classifies these assessments, and we leverage the classification outcomes to derive insights for policy-making.**

***Index Terms*—Text analysis, PU learning, Economy watchers survey**

I. Introduction

The *Economy Watcher Survey* is a market survey published by the Japanese government. The data consists of two types of sentences about assessments of economic conditions, *current and future economic conditions*, each categorized into five ranks. Although this survey provides policymakers with deep insights, it is unclear which events influence people's assessments. Clarifying the events could enable policymakers to extract more meaningful information from the survey.

This study focuses on the assessments of future economic conditions and investigate whether these assessments are based on current or future events. We assume that the assessments of the *current economic condition* are based on current events. Under this assumption, we classify people's assessments of future economic conditions into those based on *current events* and those based on *future events*, respectively. This approach leads us to propose a novel method that utilizes text data to discern these underlying bases of future economic conditions using data from the *Economy Watcher Survey*.

In the classification, we *train classifiers from positive and unlabeled data* (*PU learning*), an algorithm that enables the training of a classifier using only positive and unlabeled data. The key idea of this study is to define *current events* by using assessments of current economic conditions. After establishing what constitutes current events, we define *future events* as those not involved in assessing current economic conditions.

Upon classifying future condition assessments into those based on current and future events, we calculate the average ranks for both groups. We observe a significant difference between the economic conditions related to these two types of future assessments. This disparity suggests that people's perceptions of the future may vary significantly, which is an important consideration in economic analysis. In macroeconomics, understanding how to influence people's market expectations can be crucial. Our empirical study reveals that assessments of future economic conditions, when based on anticipated economic events, are significantly influenced by economic fundamentals such as population and diplomatic relationships.

The following sections outline our problem and propose an algorithm to address it, followed by an empirical analysis.

II. Problem Setting

We consider the binary classification problem using text data. Here, we describe the dataset and the problem in detail.

A. Economy Watchers Survey

In our analysis, we use the *Economy Watchers Survey*, a dataset that contains text data and is published by the Japanese government [1]. The purpose of this survey is to capture the regional economic trends accurately.

This survey includes assessments of current and future economic conditions, with reasons for their evaluations. The future economic condition refers to the economic situation expected three months from the survey date. Respondents rate the current and future economic conditions on a five-point scale: $0, 1, 2, 3, 4$. An evaluation of 0 indicates that the condition is "worse" or expected to "get worse" compared with a previous period, while a rating of 4 suggests it is "better" or will "get better". The middle score, 2, represents a neutral assessment.

[1]The datasets are provided on the homepage of the Japanese government, https://www5.cao.go.jp/keizai3/watcher-e/index-e.html. This survey enlists the cooperation of individuals in occupations that provide insight into regional economic activities. The dataset can be downloaded from this page.

979-8-3503-7790-3/24/$31.00 ©2024 IEEE
DOI 10.1109/IIAI-AAI63651.2024.00065

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

a) Interpretation of Assessment of Future Economic Conditions: Assessments of current and future economic conditions provide deep insights into economic realities. However, the survey does not specify the temporal context for these assessments, leading to variations in the basis for these evaluations. For example, while one respondent may base their future economic assessment on an event that occurred today, another might base theirs on an event expected in a month. Thus, to analyze these assessments more effectively, it is crucial to categorize future economic condition assessments by the timing of the events they reference.

B. Classification of Assessment of Future Economic Conditions

We aim to classify future economic condition assessments into two groups: those assessed based on current events and those based on anticipated future events. We hypothesize that assessments of current economic conditions share similarities with future assessments that are based on future events.

Our classification problem defines *assessments based on current events as positive data and those based on future events as negative data*. We assume that all assessments on the current economic conditions are positive data, while those on the future economic conditions are treated as unlabeled data, which may include both positive and negative assessments. This distinction arises from the observation that while current economic conditions are assessed based on immediate events, future conditions may be evaluated based on both current and anticipated future events. For example, assessments of economic conditions in December might be based on the current conditions in October or on anticipated events at year-end. This relationship is illustrated in Figure 1.

We train our classifier using only positive and unlabeled data, employing a PU learning algorithm. The objective is to classify $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$ into one of the two classes $\{-1, +1\}$, where $+1$ represents assessments based on current events (positive data), and -1 represents those based on future events (negative data).

C. Data-Generating Process (DGP) in the Survey

We define the DGP as follows: let there be n data points, with the i-th text data denoted as $\boldsymbol{x}_i \in \mathcal{X} \subset \mathbb{R}^d$. If text data $\boldsymbol{x}_i$ pertains to current economic events, it receives a positive label, i.e., $y_i = +1$. Conversely, if it relates to future economic events, it receives a negative label, i.e., $y_i = -1$.

In the dataset, we can observe only positive data and unlabeled data, which includes both positive and negative data. This is because while assessments of current economic conditions are assumed to be based on current events, those of future economic conditions are assumed to be based on both current and future events. Furthermore, if the text data $\boldsymbol{x}_i$ belongs to a period $t \in \{1, \dots, T\}$, we denote this fact as $z_i = t$. Using these notations, we define our DGP as

$$\{\boldsymbol{x}_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{x}|y = +1, z = t),$$
$$\{\boldsymbol{x}'_i\}_{i=1}^{n'} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{x}|z = t),$$

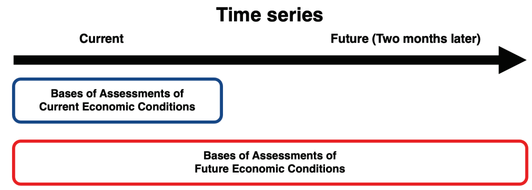

Fig. 1: Definition of the time structure of assessments.

where $\{\boldsymbol{x}_i\}_{i=1}^{n}$ and $\{\boldsymbol{x}'_i\}_{i=1}^{n'}$ denote the positive and negative data at t-th period, and $p(\boldsymbol{x}|z = t)$ can be decomposed as

$$p(\boldsymbol{x}|z = t) = p(y = +1|z = t)p(\boldsymbol{x}|y = +1, z = t) + p(y = -1|z = t)p(\boldsymbol{x}|y = -1, z = t).$$

III. PU Learning with Time Series Data

To classify data consisting only of positive and unlabeled data, we propose using *multi-task PU learning* (*MTPU*). In this section, we provide details of the proposed algorithm.

A. Learning from Positive and Unlabeled Data

Before explaining our model, let us describe the standard setting of PU learning. In PU learning, we engage in a binary classification problem to classify $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$ into one of two classes, $\{-1, +1\}$. We assume that there exists a joint density $p(\boldsymbol{x}, y)$, where $y \in \{-1, +1\}$ represents the class label of $\boldsymbol{x}$.

Let $\ell : \mathbb{R} \times \{\pm 1\} \to \mathbb{R}^+$ be a loss function, where $\mathbb{R}^+$ is the set of non-negative real values, and $\mathcal{F}$ be the set of measurable functions from $\mathcal{X}$ to $[\epsilon, 1 - \epsilon]$, where $\epsilon \in (0, 1/2)$ is a small positive value. This constant ϵ is introduced to ensure that the following optimization problem is well-defined, based on the results of Kato et al. (2019). Here, du Plessis et al. (2015a) show that the classification risk of $f \in \mathcal{F}$ can be expressed as:

$$R_{\mathrm{PU}}(f) = p(y = +1)\mathbb{E}_{\mathrm{p}}[\ell(f(X), +1)] - p(y = -1)\mathbb{E}_{\mathrm{p}}[\ell(f(X), -1)] + \mathbb{E}_{\mathrm{u}}[\ell(f(X), -1)], \quad (1)$$

where $\mathbb{E}_{\mathrm{p}}$ and $\mathbb{E}_{\mathrm{u}}$ are the expectations over $p(\boldsymbol{x}|y = +1)$ and $p(\boldsymbol{x})$, respectively. The above formulation of PU learning provides the unbiased risk of the classification problem.Z

Remark III.1 (Definition of Positive and Unlabeled Data). One of the core ideas of this study is a definition of positive and unlabeled data. In PU learning, we can train a classifier only using positive data and data with unknown labels, which implies that we can train a classifier that classifies positive and non-positive data. Thus, this classification implies that the definition of negative data is *non-positive*. This idea is helpful in a situation where it is difficult to give a clear definition of negative data. In this study, current economic condition assessments are clearly based on current economic events (positive). Therefore, we define negative data as events that are not used for reasons of current economic condition assessments. This idea is highly versatile; it can be used in many situations.

Remark III.2 (Sampling schemes in PU learning). PU learning relies on two distinct sampling schemes, namely the censoring scenario and case-control scenario (Elkan and Noto, 2008). Our

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

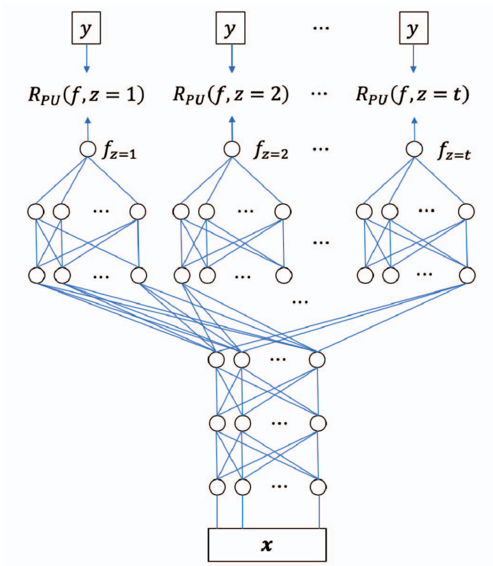

Fig. 2: Neural network model for multi-task learning for PU learning. The models share one shared network with 3 layers.

PU learning framework is the case-control scenario in which we suppose access to a positive dataset $\{\boldsymbol{x}_i\}_{i=1}^{n} \stackrel{\text{i.i.d.}}{\sim} p(\boldsymbol{x}|y=+1)$ and an unlabeled dataset $\{\boldsymbol{x}'_i\}_{i=1}^{n'} \stackrel{\text{i.i.d.}}{\sim} p(\boldsymbol{x})$.

B. Multi-Task Non-negative PU learning for Time Series Data

In addition to the standard PU learning, we could take the time structure into account. The survey comprises monthly data, with approximately $2,600$ records for each month. Here, we would need to use different classifiers for the data included in each month for the following two reasons. First, the model can vary across periods. Second, we would not be able to include data of the $(t+1)$-th period to train a model of data of the t-th period because the data of the $(t+1)$-th period might have information of the data of the t-th period. This setting makes it necessary to use different models across different periods.

For $z = t$, we denote the model as $f_{z=t}$ and the risk as

$$R_{\mathrm{PU}}(f_{z=t}, z=t) = p(y=+1|z=t)\mathbb{E}_{\mathrm{p,t}}[\ell(f(X), +1)] \\ - p(y=-1|z=t)\widehat{\mathbb{E}}_{\mathrm{p,t}}[\ell(f(X), -1)] + \widehat{\mathbb{E}}_{\mathrm{u,t}}[\ell(f(X), -1)].$$

We additionally introduce a model for *multi-task learning* to PU learning. Multi-task learning is proposed to train neural networks efficiently by utilizing common features across different tasks (Caruana, 1997). If a common feature exists across periods, we can train our models more efficiently by sharing this common feature among the models $f_{z=t}$ for $t = 1, \ldots, T$ through layers named *shared layers*. The structure of these layers is illustrated in Figure 2. We named this model *MTPU*. Details of its structure are provided in Section IV.

C. Sample Approximation of the Unbiased Risk

When training a classifier, we can naively replace the expectations with the corresponding sample averages. However, Kiryo et al. (2017a) point out that the basic form of unbiased PU learning is ineffective with deep neural networks due to over-fitting caused by the fact that the risk is not lower-bounded. To implement PU learning with deep neural networks, we apply the non-negative risk correction used in Kiryo et al. (2017a) and Kato and Teshima (2021a) to the empirical risk defined in (2). For a hypothesis set $\mathcal{H}$, we define the following risk minimization problem:

$$\hat{f}_{z=t} = \mathrm{argmin}_{f_{z=t}\in\mathcal{H}} \left[\widehat{R}_{\mathrm{PU}}(f_{z=t}, z=t) + \mathcal{R}(f)\right], \quad (2)$$

where $\widehat{R}_{\mathrm{nnPU}}(f_{z=t}, z = t)$ is a sample approximation of $R_{\mathrm{PU}}(f_{z=t}, z=t)$ with non-negative transformation proposed by Kiryo et al. (2017a) and $\mathcal{R}$ is a regularization term.

D. Class Prior and Selection Bias

The remaining problem is to make a decision regarding the class prior $p(y=+1|z=t)$. The class prior $p(y=+1|z=t)$ would be different across periods t. Although several algorithms have been proposed to estimate the class prior (du Plessis and Sugiyama, 2014; Jain et al., 2016; Ramaswamy et al., 2016), the estimation is still known to be a difficult task.

However, we can avoid the problematic estimation in the case of the particular goal we hope to reach. In our empirical analysis, we assume that the class prior is $p(y = +1|z = t) = 0.2$ for all periods, $t = 1, 2..., T$. This assumption is not realistic because the probability would have different values across the periods. However, (Kato et al., 2018) and (Kato et al., 2019) show that the function $f_{z=t}$ is simply linear-proportional to the class prior, i.e., the following relationship holds even if we miss-specify the class prior for $\boldsymbol{x}, \widetilde{\boldsymbol{x}} \in \mathcal{X}$:

$$p(y=+1|\boldsymbol{x}, z=t) \leq p(y=+1|\widetilde{\boldsymbol{x}}, z=t) \\ \Leftrightarrow f_{z=t}(\boldsymbol{x}) \leq f_{z=t}(\widetilde{\boldsymbol{x}}). \quad (3)$$

Therefore, even when we cannot obtain the exact value of $p(y=+1|\boldsymbol{x}, z=t)$, we can still identify the order of $p(y = +1|\boldsymbol{x}, z=t)$ with regard to $\boldsymbol{x}$.

Our empirical analysis classifies the assessment of future economic conditions into those based on current and future economic events. We classify $1/5$ of data from the highest value of $f_{z=t}$ into assessments based on current economic events and $1/5$ of data from the lowest value of $f_{z=t}$ into assessments based on future economic events. In addition to the robustness of the miss-specified class prior, the function $f_{z=t}$ also holds the relationship 3 under the selection bias of positive data (Kato et al., 2019) if our assumption is mild. Thus, our results can reduce the influence of the miss-specified class prior and selection bias.

IV. Empirical Study

This section reports the results of the empirical study using the Economy Watcher Survey. The survey has been conducted monthly since 2000. Our analysis utilizes data from January 2016 to January 2022, covering 72 months. Each month's data set includes approximately $3,000$ samples.

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Averaged assessments for each period and each type of economic condition are reported. For assessments based on current and future events, we conducted a two-sample t-test. A significant difference between the mean values of the assessments is indicated by a superscript $*$. One $*$ denotes that the null hypothesis of the two-sample t-test is rejected at the 5% significance level, while two $**$s indicate rejection at the 1% significance level. The MTPU represents our proposed method. "The Original" refers to the original category in the Economy Watcher Survey. PU1 and PU2 are alternative existing methods used for comparison. "Current" and "Future" are original categories, while "Classified Current" and "Classified Future" are categories classified into current and future assessments by PU learning. We report the averaged assessments for each category.

		MTPU		Original		PU1		PU2	
		Classified Current	Classified Future	Current	Future	Classified Current	Classified Future	Classified Current	Classified Future
Jan.	2016	1.842**	1.925**	1.996	1.864	1.889	1.857	1.846	1.964
Feb.	2016	1.701**	1.720**	1.949	1.780	1.749	1.756	1.733	1.876
Mar.	2016	1.768*	1.694*	1.870	1.800	1.805	1.731	1.691	1.804
Apr.	2016	1.606**	1.668**	1.822	1.719	1.764	1.625	1.717	1.684
May	2016	1.574**	1.736**	1.896	1.683	1.661	1.680	1.582	1.808
June	2016	1.588	1.622	1.656	1.632	1.671	1.531	1.588	1.685
July	2016	1.721**	1.784**	1.887	1.804	1.797	1.864	1.725	1.860
Aug.	2016	1.770**	1.906**	1.904	1.815	1.789	1.863	1.695	1.887
Sept.	2016	1.661**	1.860**	1.961	1.781	1.700	1.767	1.650	1.887
Oct.	2016	1.789**	1.914**	1.985	1.852	1.785	1.878	1.762	1.906
Nov.	2016	2.028	1.865	1.978	1.960	1.936	1.956	1.944	1.948
Dec.	2016	2.139**	1.975**	1.974	2.095	2.053	2.074	2.151	1.996
Jan.	2017	2.036	1.888	1.997	1.959	1.897	1.932	2.008	2.008
Feb.	2017	1.947**	1.992**	2.091	1.969	1.886	2.024	1.886	2.033
Mar.	2017	2.181*	1.881*	1.967	2.040	2.122	1.984	2.157	1.968
Apr.	2017	2.176	1.963	2.034	2.039	2.049	1.959	2.135	1.955
May	2017	2.077*	1.963*	2.080	2.013	1.988	2.041	2.061	2.008
June	2017	2.086*	1.939*	2.084	2.022	1.951	1.971	2.016	2.078
July	2017	2.162	1.980	2.034	2.055	2.077	2.033	2.016	1.972
Aug.	2017	2.024	1.904	2.011	1.975	1.996	1.952	2.008	1.956
Sept.	2017	2.041	1.914	2.032	2.005	1.943	1.967	1.931	2.033
Oct.	2017	1.909**	2.040**	2.175	1.998	1.822	2.048	1.905	2.139
Nov.	2017	2.278	2.045	2.076	2.125	2.173	2.020	2.121	2.061
Dec.	2017	2.140**	2.137**	2.067	2.171	2.108	2.068	2.240	2.177
Jan.	2018	1.935**	1.951**	2.124	1.957	1.874	2.016	1.935	2.029
Feb.	2018	1.831**	1.938**	2.125	1.936	1.778	1.942	1.926	1.950
Mar.	2018	2.052*	2.032*	2.005	2.073	2.105	1.943	2.073	1.992
Apr.	2018	2.184	1.922	2.071	2.069	2.123	1.984	2.115	2.057
May	2018	1.907**	1.837**	2.039	1.900	1.870	1.894	1.878	1.963
June	2018	1.860**	1.963**	2.034	1.910	1.848	1.942	1.835	2.021
July	2018	1.883**	1.838**	1.965	1.874	1.785	1.887	1.895	1.919
Aug.	2018	1.900**	1.988**	2.034	1.913	1.799	1.931	1.956	2.016
Sept.	2018	1.735**	1.881**	2.041	1.882	1.861	1.918	1.682	2.000
Oct.	2018	1.964**	1.834**	2.006	1.916	1.948	1.785	1.911	1.887
Nov.	2018	2.049	1.909	2.033	1.980	1.943	1.930	1.947	1.979
Dec.	2018	2.017*	1.862*	1.881	1.951	1.996	1.900	2.017	1.912
Jan.	2016	1.780**	1.784**	2.016	1.800	1.748	1.833	1.756	1.833
Feb.	2019	1.871**	1.925**	2.014	1.863	1.917	1.837	1.829	1.95
Mar.	2019	1.975*	1.856*	1.943	1.879	2.004	1.797	1.895	1.856
Apr.	2019	1.852	1.894	1.972	1.916	1.877	1.962	1.797	1.928
May	2019	1.845**	1.762**	1.863	1.766	1.853	1.713	1.784	1.758
June	2019	1.736**	1.676**	1.868	1.719	1.762	1.718	1.715	1.748
July	2019	1.544**	1.684**	1.744	1.632	1.552	1.696	1.560	1.773
Aug.	2019	1.669**	1.690**	1.526	1.677	1.616	1.673	1.645	1.682
Sept.	2019	1.939**	1.744**	1.426	1.850	2.065	1.671	2.073	1.71
Oct.	2019	1.221**	1.641**	1.737	1.413	1.257	1.492	1.269	1.589
Nov.	2019	1.544**	1.629**	1.823	1.530	1.519	1.621	1.456	1.638
Dec.	2019	1.604**	1.674**	1.782	1.617	1.658	1.669	1.633	1.640

We employ the Bag-of-Words model to represent the documents as $20,022$-dimensional vectors. After vectorization, we apply PU learning using the multi-task PU learning (MTPU) method. Additionally, we apply a standard model of PU learning for performance comparison. This method is implemented in two ways: first, using all samples to train a single model; second, preparing a separate model for each month. Further details on the neural networks are provided in the subsequent section. After training, we classify assessments of future economic conditions using the unlabeled data involved in training.

a) Neural network model: First, we describe the model used for MTPU. The shared network portion of the model is a 3-layer *multilayer perceptron* (MLP) with ReLU activation (Nair and Hinton, 2010), specifically configured as 16914-500-500-500. The neural network model following the shared network is a 2-layer MLP, specifically 500-500-1, also with ReLU activation. Next, we discuss the model used for non-negative PU learning, which is a 5-layer MLP (configured as 16914-500-500-500-500-1) with ReLU. We set $p(y = +1|z = t) = 0.2$ for all periods $t \in \{1, 2, ..., 72\}$. For both methods, we use logistic loss as the loss function.

A. *Difference among the Assessments*

This section reports the extent to which assessments differ across current and future economic condition assessments and

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

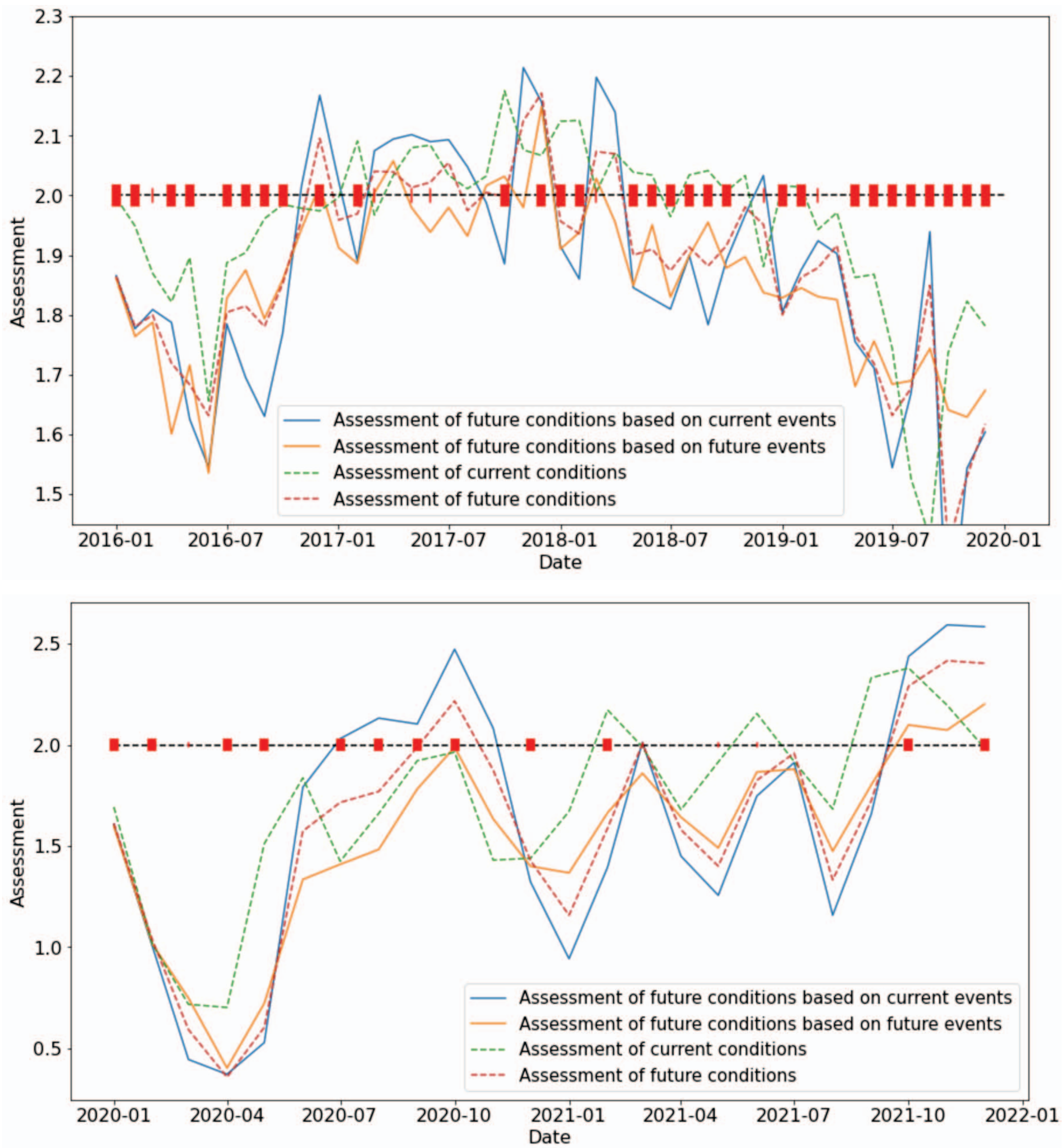

Fig. 3: Plotted assessments of the economic conditions based on current and future events. The horizontal line at $y = 2$ is the neutral state. The red vertical lines on the horizontal line represent the results of the two-sample t-test. The thin and bold red vertical lines represent the 5% and 1% significance levels, respectively.

future economic conditions based on current and future events.

a) Averaged Assessments and t-test: We report the average assessments of future economic conditions based on current and future events in comparison with those of the entire set of current and future economic condition assessments. Future economic condition assessments are classified into those based on current and future economic events by MTPU and non-negative PU learning with neural networks. For non-negative PU learning, we use two models. The first model (named PU1) involves training one model for all samples. The second model (named PU2) uses different models for the data of different months.

The results are presented in Table I. We only display data from January 2016 to June 2019 due to page limitations. Additionally, the results after this period are significantly influenced by the COVID-19 pandemic, which complicates the interpretation of the data. For each period, we show the results of the two-sample t-test with unequal variances between the assessments of economic conditions based on current and future events. Values where the difference between the means of the assessments is significant are indicated by a superscript $*$ in the table. One $*$ and two $*$s indicate that the null hypothesis of the two-sample t-test is rejected at the 5% and 1% significance levels, respectively.

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

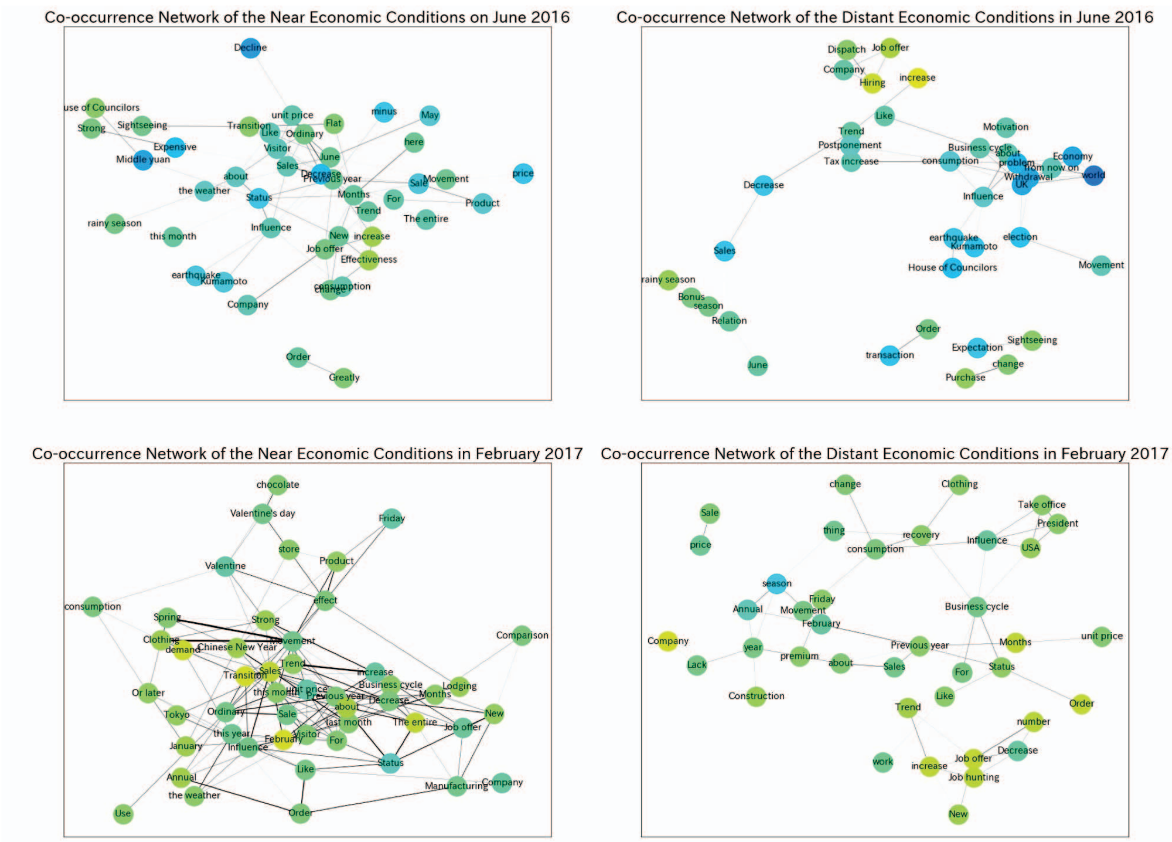

Fig. 4: Co-occurrence network of assessments based on current and future events in June 2016 and February 2017. The lengths of the edges represent the value of the Jaccard coefficients. Shorter edges indicate a stronger relationship (the value of the Jaccard coefficients is larger) between the two words. The widths of the edges also represent the value of the Jaccard coefficients between the two words. The bold edges similarly signify a stronger relationship (the value of the Jaccard coefficients is larger) between the two words. The color of the nodes relates to the assessment. The yellow-green color denotes that the average value is 2, i.e., the assessment is neutral. The warmer and cooler colors represent positive and negative assessments, respectively.

This result suggests that although the original data consolidate future economic assessments into a single category, these are potentially composed of two distinct evaluations. Specifically, there are assessments of future economic conditions based on events in the near term from the current point in time, and assessments interested in events further in the future.

b) Visualization as a time series: To facilitate a more intuitive understanding of the reported results, we plotted the averaged assessments as a time series in Figure 3, where the x-axis represents the time series, and the y-axis represents the value of the assessment. The blue, orange, and red lines correspond to assessments of future economic conditions. The red line represents the averaged assessments in the original data. Our results are shown by the blue and orange lines: the blue line represents assessments based on current events (positive data), and the orange line represents assessments based on future events (negative data). The green line corresponds to assessments of current economic conditions. The horizontal black dashed line at $y = 2.0$ represents the neutral condition in the 5-step evaluations of economic conditions from 0 (bad) to 4 (good). The vertical red lines perpendicular to the line $y = 2.0$ indicate that the difference between the average assessments of economic conditions based on current and future events is significant in the two-sample t-test. Bold vertical lines signify that the null hypothesis of the two-sample t-test is rejected at the 1% significance level, and thinner red lines indicate a rejection at the 5% level. For example, the assessments of the future economic conditions based on current events in 2017 are significantly higher than those based on future events.

This result visually demonstrates that future economic assessments in the original data can be divided into assessments based on events close to the current point in time and those interested in events further in the future. From this figure, critical information can be extracted for economic practitioners. For example, even if the assessments of future conditions based on current events (represented by the line of blue dots) are high, it is sometimes observed that assessments of future conditions based on future events (represented by the line of green dots)

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

are low. This indicates that people tend to have a negative view of the economy with respect to more long-term factors. Conversely, the opposite can also occur. Particularly, when the values of both assessments are significantly divergent, it suggests that economic judgments vary greatly depending on the events being considered, implying that policymakers need to evaluate people's economic sentiments more carefully.

B. Co-occurrence Networks

This section presents our analysis of the text data based on assessments of future economic conditions. For text mining, we employ *tf-idf* and the *Jaccard coefficient*, which are standard techniques in natural language processing.

First, we classify the assessments of future economic conditions into those based on current and future events for the month in which the assessments were made, thereby forming groups of monthly assessments. We denote the set of these groups as $\mathcal{M}$, and apply tf-idf to identify words that characterize each document. For the top 50 words with the highest tf-idf scores, we measure the Jaccard coefficient (Manning and Schütze, 1999), which quantifies the similarity between two sets. Let $\mathcal{M}_w \in \mathcal{M}$ be a set of sentences including the word w. The Jaccard coefficient $J(\mathcal{M}_a, \mathcal{M}_b)$ for words a and b is given by

$$J(\mathcal{M}_a, \mathcal{M}_b) = \frac{|\mathcal{M}_a \cap \mathcal{M}_b|}{|\mathcal{M}_a \cup \mathcal{M}_b|}. \quad (4)$$

Based on these results, we plotted the co-occurrence networks in Figure 4.[2] Due to space limitations, we only display the network of assessments for June 2016 and February 2017. June 2016 was a period during which the value of assessments changed significantly. Throughout 2017, the assessments of future economic conditions based on current events were fewer than those based on future events, with February 2017 being one of these periods.

C. Interpretations

Figure 4 displays words related to economic fundamentals, such as the structure of the labor supply and international politics. This suggests that assessments of future economic conditions based on current events are indicative of myopic viewpoints, while assessments based on future events reflect long-term perspectives.

For instance, words such as "UK" and "withdrawal," which relate to Brexit, appear in the context of future economic conditions based on current events from June 2016, alongside terms like "Business cycle" and "Trend." Conversely, in February 2017, terms like "US" and "President" emerged. On the other hand, the economic conditions based on future events in June 2016 and February 2017 are characterized by terms less associated with economic fundamentals, such as "rainy season" and "Valentine's day." For policymakers, this finding is insightful because it implies that altering people's expectations may be challenging if their assessments are grounded in economic fundamentals.

[2]We translated from Japanese to English using an API provided by Google (https://pypi.org/project/googletrans/).

V. Related Work

Text analysis has garnered attention as a method for studying economic trends. Pioneering methods include those of Tetlock (2007) and Tetlock et al. (2008), which involve constructing sentiment indexes from articles in the *Wall Street Journal* and analyzing market predictability. Kulkarni et al. (2009) predicts residential prices using the number of searches on Google, while Guzman (2011) constructs real-time inflation expectations from Google search queries. Algaba et al. (2020) discusses the integration of econometrics and sentiment analysis. Shapiro et al. (2022) introduces a new time-series measure of economic sentiment.

PU learning is distinguished by two sampling schemes: *one-sample* and *two-sample* scenarios (Elkan and Noto, 2008; Niu et al., 2016), where the former is more general (Niu et al., 2016). In the two-sample scenario, du Plessis et al. (2015b) develops an unbiased risk minimization framework, known as *unbiased PU learning*. Kiryo et al. (2017b) applies a non-negative correction to prevent neural networks from overfitting in this context. PU learning methods are closely related to other problem settings such as density-ratio estimation (Kato and Teshima, 2021b) and semi-supervised classification (Sakai et al., 2017, 2018). PU learning under selection bias has also been extensively studied (Bekker and Davis, 2018; Bekker et al., 2019; Hsieh et al., 2019; Kato et al., 2019; Luo et al., 2021; Marlin and Zemel, 2009; Schnabel et al., 2016).

Li and Liu (2003), Yu and Li (2007), and Lu and Bai (2010) apply PU learning methods for text classification using formulations different from ours. Independently, Charoenphakdee et al. (2019) and Jacovi et al. (2021) have applied methods based on unbiased PU learning, albeit with different aims, methods, and empirical analyses.

VI. Conclusion

In this study, we proposed an application of PU learning and text mining to financial text data. We developed a new model named MTPU to train neural networks efficiently using data with a temporal structure. Our empirical analysis demonstrated the classification results and interpretations based on text mining and economic principles. These results are insightful for policymakers, suggesting that individuals may base their assessments of future economic conditions on different events, potentially leading to varied economic outlooks.

References

Andres Algaba, David Ardia, Keven Bluteau, Samuel Borms, and Kris Boudt. 2020. Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*, 34(3):512–547.

Jessa Bekker and Jesse Davis. 2018. Learning from positive and unlabeled data under the selected at random assumption. In *Proceedings of the Second International Workshop on Learning with Imbalanced Domains: Theory and Applications*, volume 94, pages 8–22.

Jessa Bekker, Pieter Robberechts, and Jesse Davis. 2019. Beyond the selected completely at random assumption for

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.

learning from positive and unlabeled data. In *KDD*, page 71–85.

Rich Caruana. 1997. Multitask learning. *Machine Learning*, 28(1):41–75.

Nontawat Charoenphakdee, Jongyeong Lee, Yiping Jin, Dittaya Wanvarie, and Masashi Sugiyama. 2019. Learning only from relevant keywords and unlabeled documents. In *Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3993–4002.

Marthinus Christoffel du Plessis, Gang. Niu, and Masashi Sugiyama. 2015a. Convex formulation for learning from positive and unlabeled data. In *ICML*, pages 1386–1394.

Marthinus Christoffel du Plessis, Gang. Niu, and Masashi Sugiyama. 2015b. Convex formulation for learning from positive and unlabeled data. In *ICML*, pages 1386–1394.

Marthinus Christoffel du Plessis and Masashi Sugiyama. 2014. Class prior estimation from positive and unlabeled data. *IEICE Transactions on Information and Systems*, E97-D(5):1358–1362.

Charles Elkan and Keith Noto. 2008. Learning classifiers from only positive and unlabeled data. In *ICDM*, pages 213–220.

Giselle Guzman. 2011. Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of Economic and Social Measurement*, 36.

Yu-Guan Hsieh, Gang Niu, and Masashi Sugiyama. 2019. Classification from positive, unlabeled and biased negative data. In *ICML*, volume 97, pages 2820–2829.

Alon Jacovi, Gang Niu, Yoav Goldberg, and Masashi Sugiyama. 2021. Scalable evaluation and improvement of document set expansion via neural positive-unlabeled learning. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 581–592.

Shantanu Jain, Martha White, Michael W Trosset, and Predrag Radivojac. 2016. Nonparametric semi-supervised learning of class proportions. In *NIPS*.

Masahiro Kato and Takeshi Teshima. 2021a. Non-negative bregman divergence minimization for deep direct density ratio estimation. In *International Conference on Machine Learning (ICML)*, Proceedings of Machine Learning Research, pages 5320–5333. PMLR.

Masahiro Kato and Takeshi Teshima. 2021b. Non-negative bregman divergence minimization for deep direct density ratio estimation. In *ICLR*, volume 139, pages 5320–5333.

Masahiro Kato, Takeshi Teshima, and Junya Honda. 2019. Learning from positive and unlabeled data with a selection bias. In *ICLR*.

Masahiro Kato, Liyuan Xu, Gang Niu, and Masashi Sugiyama. 2018. Alternate estimation of a classifier and the class-prior from positive and unlabeled data. *arXiv*.

Ryuichi Kiryo, Gang Niu, Marthinus Christoffel du Plessis, and Masashi Sugiyama. 2017a. Positive-unlabeled learning with non-negative risk estimator. In *NIPS*, pages 1675–1685.

Ryuichi Kiryo, Gang Niu, Marthinus Christoffel du Plessis, and Masashi Sugiyama. 2017b. Positive-unlabeled learning with non-negative risk estimator. In *NeurIPS*, pages 1675–1685.

Rajendra Kulkarni, Kingsley Haynes, Roger Stough, and Jean Paelinck. 2009. Forecasting housing prices with google econometrics. *SSRN Electronic Journal*.

Xiaoli Li and Bing Liu. 2003. Learning to classify texts using positive and unlabeled data. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, page 587–592. Morgan Kaufmann Publishers Inc.

Fang Lu and Qingyuan Bai. 2010. Semi-supervised text categorization with only a few positive and unlabeled documents. In *2010 3rd International Conference on Biomedical Engineering and Informatics*, volume 7, pages 3075–3079.

Chuan Luo, Pu Zhao, Chen Chen, Bo Qiao, Chao Du, Hongyu Zhang, Wei Wu, Shaowei Cai, Bing He, Saravanakumar Rajmohan, and Qingwei Lin. 2021. Pulns: Positive-unlabeled learning with effective negative sample selector. *AAAI*, 35(10):8784–8792.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA.

Benjamin M. Marlin and Richard S. Zemel. 2009. Collaborative prediction and ranking with non-random missing data. In *RecSys*, pages 5–12.

Vinod Nair and Geoffrey E. Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *ICML*.

Gang Niu, Marthinus Christoffel du Plessis, Tomoya Sakai, Yao Ma, and Masashi Sugiyama. 2016. Theoretical comparisons of positive-unlabeled learning against positive-negative learning. In *NeurIPS*, volume 29.

Harish Ramaswamy, Clayton Scott, and Ambuj Tewari. 2016. Mixture proportion estimation via kernel embeddings of distributions. In *ICML*, pages 2052–2060.

Tomoya Sakai, Marthinus Christoffel du Plessis, Gang Niu, and Masashi Sugiyama. 2017. Semi-supervised classification based on classification from positive and unlabeled data. In *ICML*, volume 70, pages 2998–3006.

Tomoya Sakai, Gang Niu, and Masashi Sugiyama. 2018. Semi-supervised auc optimization based on positive-unlabeled learning. *Machine Learning*, 107(4):767–794.

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *ICML*, volume 48, pages 1670–1679, New York, New York, USA.

Adam Hale Shapiro, Moritz Sudhof, and Daniel J. Wilson. 2022. Measuring news sentiment. *Journal of Econometrics*, 228(2):221–243.

Paul C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Paul C. Tetlock, Maytal Saar-tsechansky, and Sofus Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3):1437–1467.

Shuang Yu and Chunping Li. 2007. Pe-puc: A graph based pu-learning approach for text classification. In *Machine Learning and Data Mining in Pattern Recognition*, pages 574–584, Berlin, Heidelberg. Springer Berlin Heidelberg.

Authorized licensed use limited to: Columbia University Libraries. Downloaded on August 18,2025 at 15:51:39 UTC from IEEE Xplore. Restrictions apply.