MEASURING ECONOMIC SENTIMENT: A TEXT MINING APPROACH

HYEYOUNG LEE, GYUTAE KIM, MOONHEE LEE, YOUNGHWAN LEE, HEEJIN HWANG

BANK OF KOREA

MOTIVATION

- Economic sentiments is deemed crucial in explaining economic activities. (see. [1])

- Despite its increasing importance, measuring the economic sentiment by sample survey is expensive and time-consuming.

- Recent surge in machine learning researches pave the way to tackle these challenges.

- The endeavors such as [3] aim at retrieving the infrequent and expensive sentiment information from frequent and publicly available source of data like news article using machine learning techniques.

- In a similar vein, we try to measure the economic sentiment from news articles in Korea based on [2] and [3].

WORK FLOW: SNAPSHOT
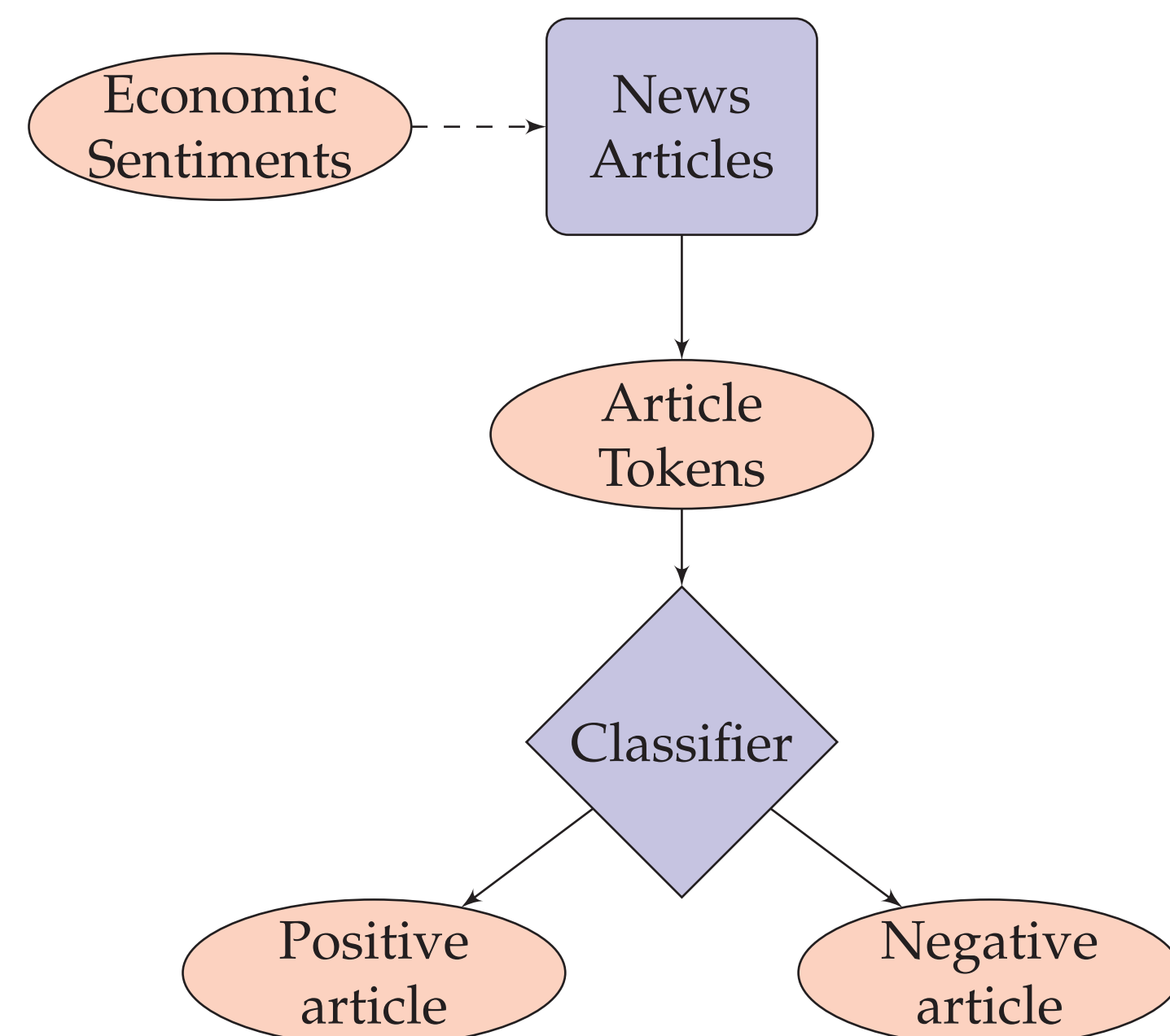
$$NSI = 100 + 100 \times \frac{\text{\# of positive articles} - \text{\# of negative articles}}{\text{\# of positive articles} + \text{\# of negative articles}}$$

- News articles collected by public institution[a] with web crawling from the Korean portal service Naver.

- Period: 1st Jan. 2008 to 31st Oct. 2018

- Size of data: approximately 9 million cases

- Comma separated value (CSV) typed database.

[a]NIA: National Information Society Agency

CLASSIFIER: MODEL APPROACH VS VOCABULARY APPROACH

Model approach

- Embed words into the real vector space using Word2Vec library.

- To construct the training set, label articles according to the economic sentiments.

- Using vectorized articles and labelled training set, fit the model to build a classifier.

- We try LR (Logistic Regression), LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), Boosting, RF (Random Forest), and NN (Neural Network). The performance of SVM classifier is the best according to the F1 score.

Vocabulary approach
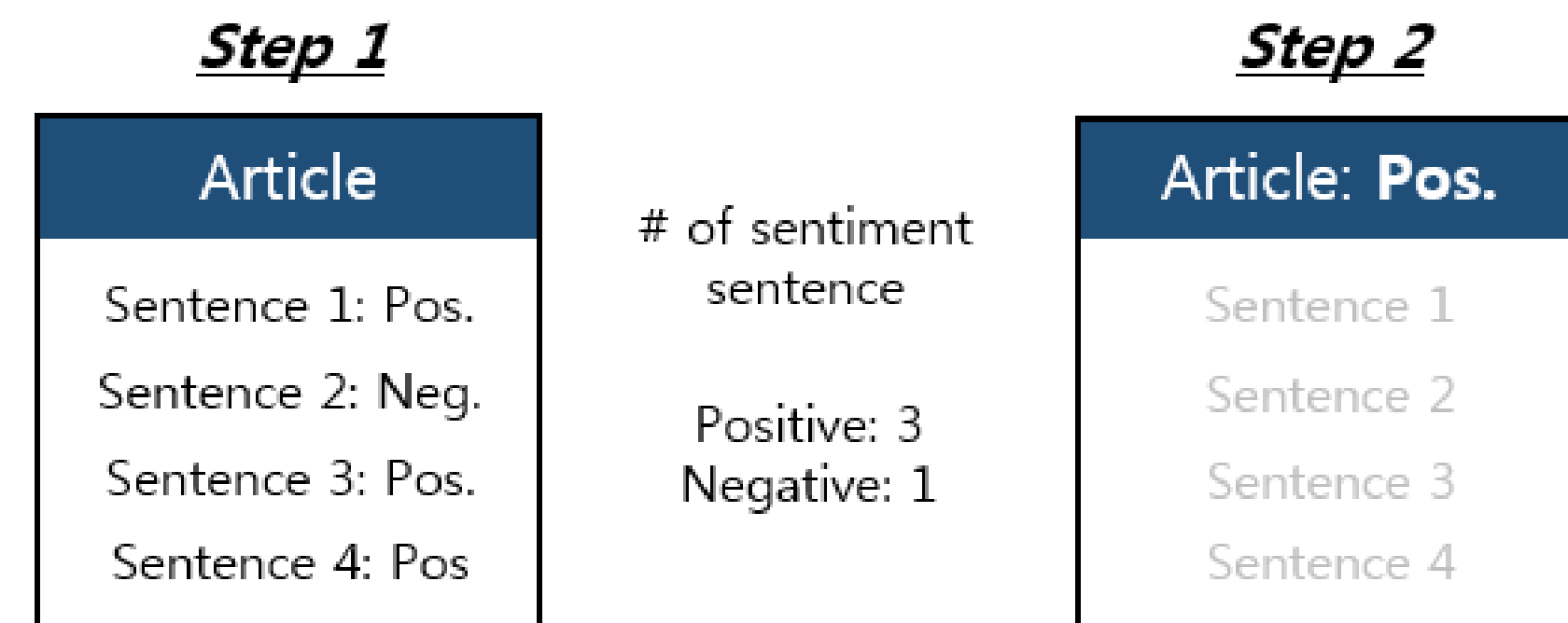
- Depend on the dictionary for economic sentiment, judge the tone of article.

PERFORMANCE: MODEL BASED CLASSIFIER

Table. Performance Evaluation (Model approach)

Model	Sentiment	Accuracy	Sensitivity	Specificity	Precision	F1 Score
LR	Positive	0.78	0.77	0.78	0.65	0.71
	Negative	0.78	0.78	0.77	0.87	0.82
LDA	Positive	0.74	0.87	0.67	0.58	0.69
	Negative	0.74	0.67	0.87	0.91	0.77
SVM	Positive	0.81	0.76	0.83	0.70	0.73
	Negative	0.81	0.83	0.76	0.87	0.85
Boosting	Positive	0.76	0.46	0.91	0.73	0.56
	Negative	0.76	0.91	0.46	0.77	0.83
RF	Positive	0.75	0.36	0.95	0.78	0.49
	Negative	0.76	0.92	0.43	0.76	0.83
NN	Positive	0.80	0.65	0.88	0.74	0.70
	Negative	0.80	0.90	0.61	0.82	0.86

		True Condition	
		Positive	Negative
Predicted Condition	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}, \text{Precision} = \frac{TP}{TP+FP}$$

$$\text{F1 Score} = \frac{2TP}{2TP+FP+FN}$$

CONCLUSION AND REMAINING ISSUES

- Reusing training samples for the calculation of NSI may cause the overfitting problem.

- In the classification, it is necessary to separate the economic agent to improve of usefulness of new index.

- In addition, we want to try the sentiment analysis based on sentences as well as news articles.

- It is necessary to consult with experts to elaborate the sentiment dictionary to improve the accuracy of classification.

NEWS SENTIMENT INDEX
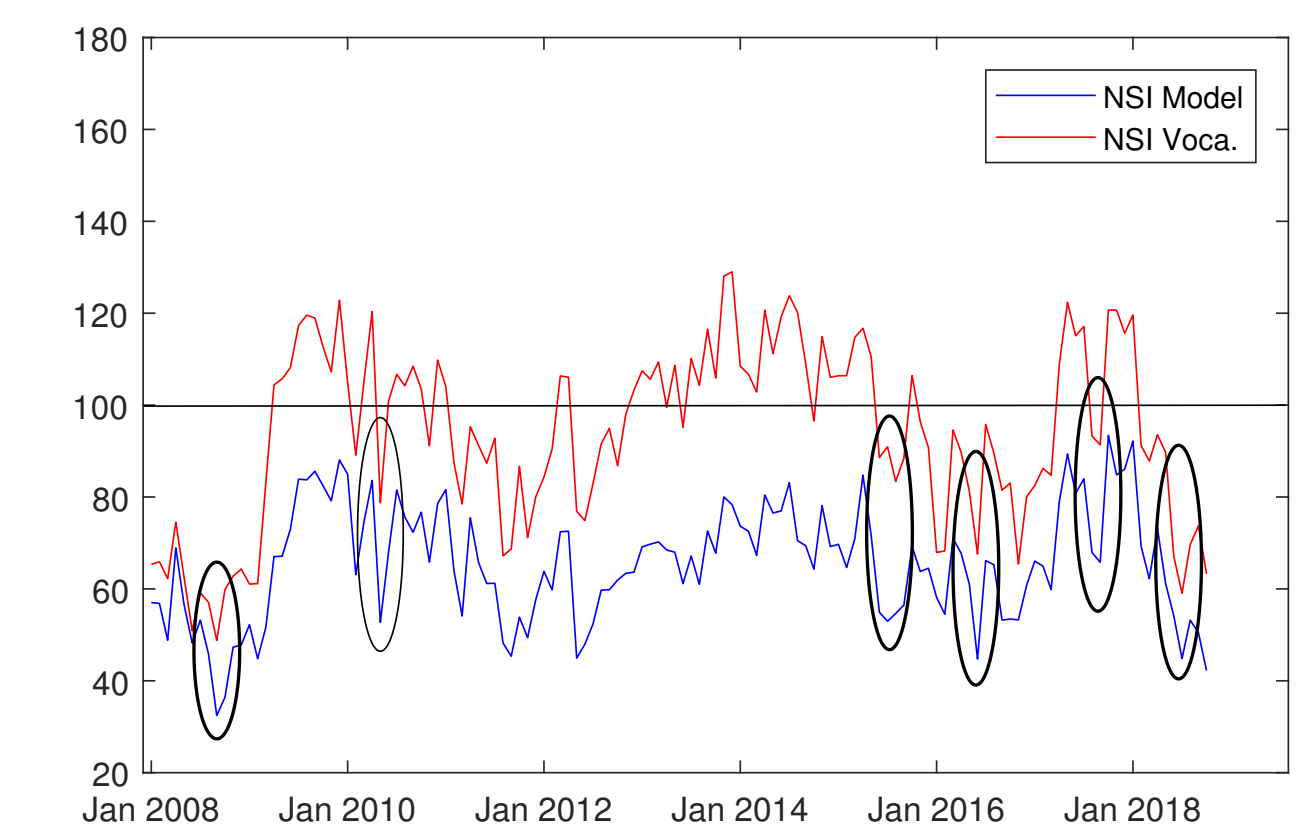
- Two of them are highly correlated; correlation coefficient is 0.896.

- Both of them reflect the major economic issues such as 2008 financial crisis and eurozone crisis.

EXTERNAL VALIDITY
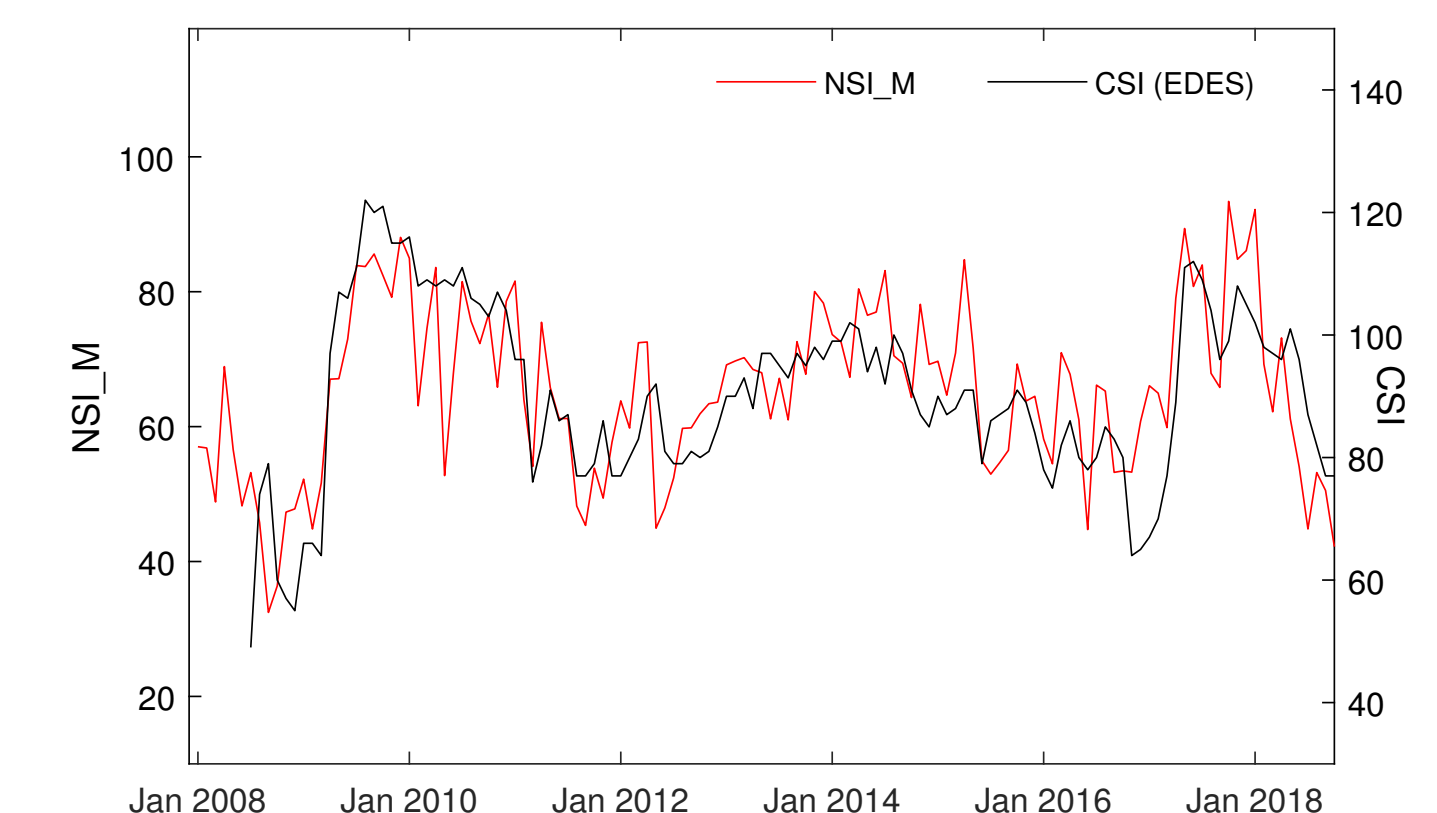
- The new index seems to have similar pattern to the survey based CSI EDES (Consumer Sentiment Index on Expectations of Domestic Economic Situation).

- Furthermore, the new index turns out to lead CSI EDES by one month according to the cross correlation analysis.

REFERENCES

[1] Jess Benhabib and Mark M Spiegel. Sentiments and economic activity: Evidence from US states. *The Economic Journal*, 129(618):715–733, 2018.

[2] Dhanush Dharmaretnam, Madhav Malhotra, Jasmeet Singh, and Evan Wilde. Sentiment analysis of news articles. *Working Paper*.

[3] Adam Hale Shapiro, Moritz Sudhof, and Daniel Wilson. Measuring news sentiment. *Federal Reserve Bank of San Francisco Working Paper Series*.