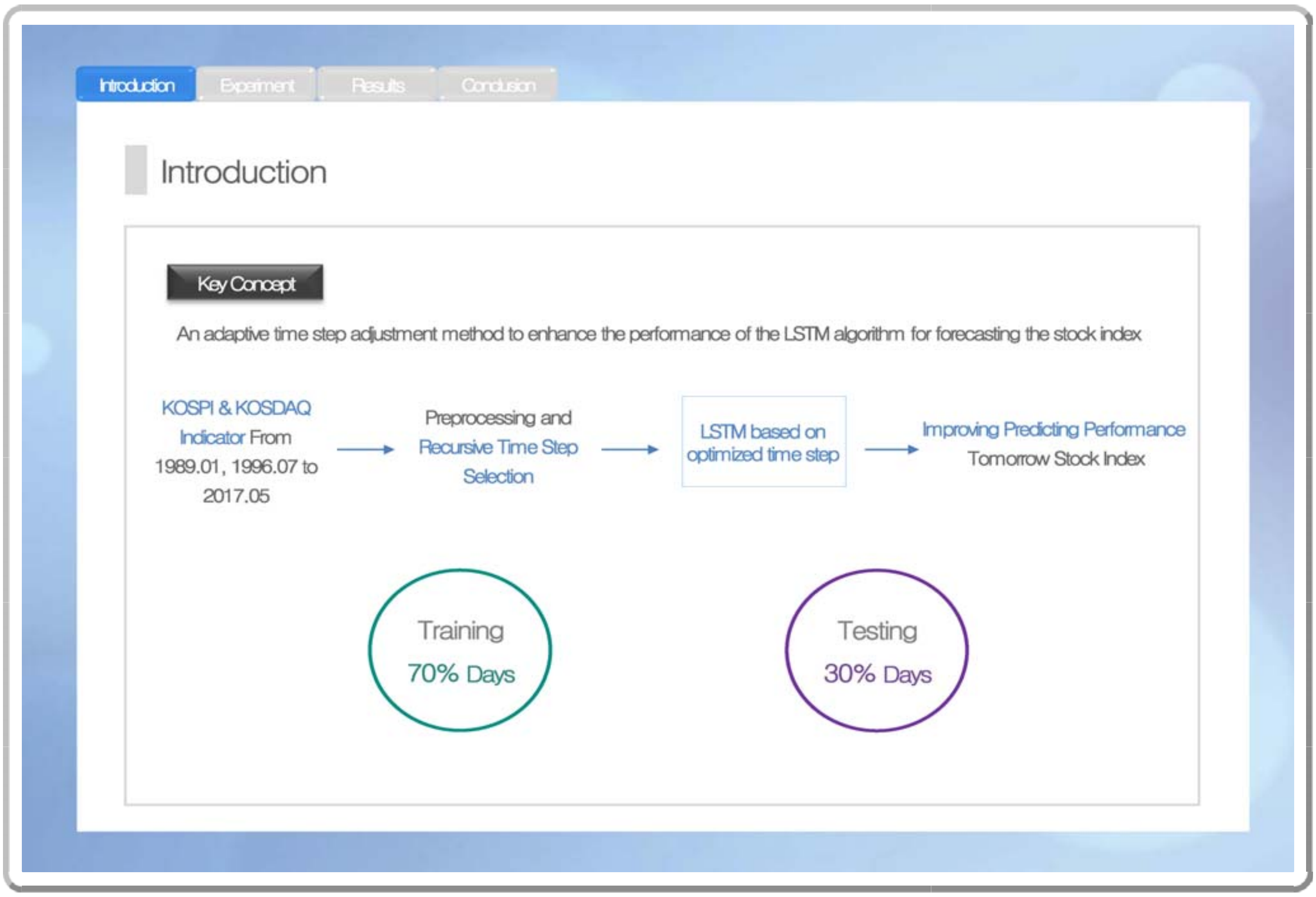
Long Short-Term Memory(LSTM)에서 시간 단위 민감도 분석을 통한 주가지수 예측


Long Short-Term Memory(LSTM)에서 시간
단위 민감도 분석을 통한 주가지수 예측
UNIST
Song Donghwan, Moise Busogi, Namhun Kim
2018. 04. 07


Introduction
Experiment
Results
Conclusion
Introduction
Key Concept
An adaptive time step adjustment method to enhance the performance of the LSTM algorithm for forecasting the stock index
KOSPI & KOSDAQ
Indicator From
1989.01, 1996.07 to
2017.05
Preprocessing and
Recursive Time Step
Selection
LSTM based on
optimized time step
Improving Predicting Performance
Tomorrow Stock Index
Training
70% Days
Testing
30% Days

Introduction | Experiment | Results | Conclusion

Dataset Description

Description

Research Data : KOSPI INDEX	Research Data : KOSDAQ INDEX
Total sample : 7457 Trading days	Total sample : 5281 Trading days
From Jan 1989 to May 2017	From July 1996 to May 2017
5220 (70%) – Training Data / 2237 (30%) – Test Data	3697 (70%) – Training Data / 1584 (30%) – Test Data
Scaled into the range of [0, 1.0] – Min/Max Scaling	Scaled into the range of [0, 1.0] – Min/Max Scaling
Time Step : [5,10, ⋯ , 115, 120]	Time Step : [5,10, ⋯ , 115, 120]
Hidden Layer : 10	Hidden Layer : 10
Iteration : 5000	Iteration : 5000
Input Features : 7 (Open, High, Low, Close, Volume, Turnover, Diff)	Input Features : 7 (Open, High, Low, Close, Volume, Turnover, Diff)
Output Feature : 1 (Tomorrow KOSPI Index)	Output Feature : 1 (Tomorrow KOSDAQ Index)

Introduction | Experiment | Results | Conclusion

Long short-term memory

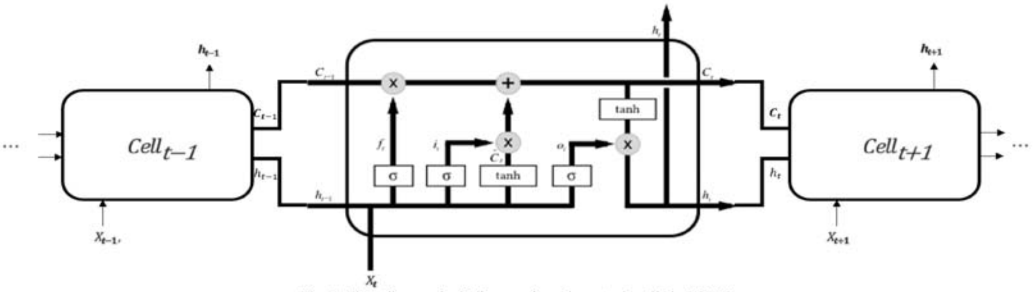


Fig. 1. Flow diagram including previous time step's cells in LSTM

$X_t = \{x \mid x_t^1, x_t^2, \cdots, x_t^k\}$: *Input vector at t*

$i_t = \sigma(W_{xi} \cdot X_t^n + W_{hi} \cdot h_{t-1} + b_i)$	: *Input gate*	$\tilde{C}_t = \tanh(W_{xc} \cdot X_t^n + W_{hc} \cdot h_{t-1} + b_c)$	: *Sub Cell*
$f_t = \sigma(W_{xf} \cdot X_t^n + W_{hf} \cdot h_{t-1} + b_f)$	: *forget gate*	$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$	: *New Cell*
$o_t = \sigma(W_{xo} \cdot X_t^n + W_{ho} \cdot h_{t-1} + b_o)$	: *output gate*	$h_t = o_t * \tanh(C_t)$	: *Output*

LSTM is a one of the most powerful algorithms for predicting timeseries data in deep learning

Introduction Experiment Results Conclusion

Comparison Models

Support Vector Regression

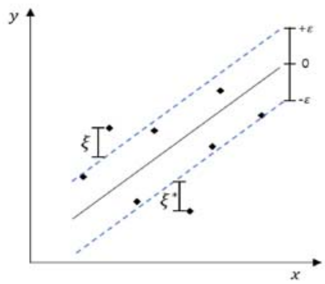


Fig. 2. Support Vector Regression with epsilon intensive band.

Artificial neural network

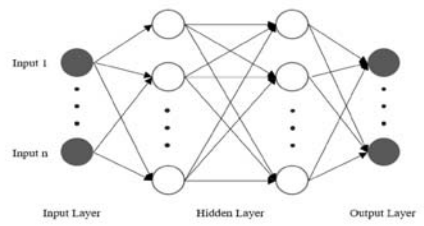


Fig. 3. Artificial neural networks model.

SVM, BPNN and Vanilla RNN are compared with t-LSTM

Introduction Experiment Results Conclusion

Performance results for each time step

Performance Measures

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - f_i}{y_i}\right|$$

Table1 Performance results of KOSPI and KOSDAQ for each time step

	KOSPI			KOSDAQ		
Time Steps	MAE	RMSE	MAPE	MAE	RMSE	MAPE
5	17.627	23.254	0.992	6.533	9.921	1.102
10	16.613	22.323	0.943	5.977	8.210	1.033
15	16.419	22.131	0.935	5.561	7.858	0.969
20	16.351	21.974	0.930	5.988	8.515	1.029
25	16.416	21.999	0.933	6.519	9.229	1.109
30	16.726	22.248	0.948	6.634	9.506	1.127
35	17.192	22.665	0.970	6.960	9.996	1.178
40	17.339	22.803	0.978	7.097	10.186	1.200
45	17.594	23.028	0.990	7.283	10.450	1.230
50	17.639	23.072	0.992	7.352	10.540	1.241
55	17.633	23.073	0.992	7.294	10.438	1.232
60	17.728	23.162	0.996	7.368	10.547	1.244
65	17.861	23.286	1.003	7.403	10.594	1.249
70	17.875	23.297	1.004	7.474	10.700	1.260
75	17.842	23.266	1.002	7.437	10.633	1.255
80	17.993	23.405	1.009	7.482	10.699	1.262
85	18.070	23.473	1.013	7.418	10.605	1.252
90	18.219	23.603	1.020	7.537	10.781	1.270
95	18.318	23.687	1.025	7.484	10.695	1.262
100	18.324	23.693	1.025	7.525	10.759	1.268
105	18.306	23.673	1.024	7.576	10.828	1.276
110	18.320	23.687	1.024	7.621	10.895	1.283
115	18.367	23.735	1.026	7.508	10.719	1.266
120	18.317	23.693	1.024	7.554	10.788	1.273

Table 2 Comparison table of t-LSTM and other algorithms

	KOSPI			KOSDAQ		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
t-LSTM	16.351	21.974	0.930	5.561	7.858	0.969
RNN	17.084	22.500	0.963	6.431	8.563	1.116
BPN	64.790	22.834	0.940	5.530	9.973	3.730
SVR	53.026	59.008	2.812	31.977	35.651	5.682

Introduction | Experiment | **Results** | Conclusion

Error Graph for each time steps

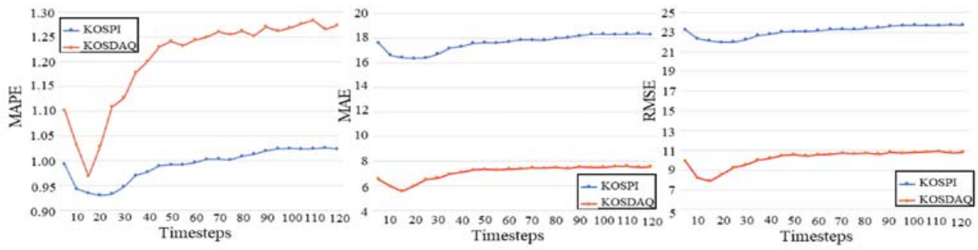


Fig. 4. MAPE, MAE and RMSE measured at each time step in the prediction of KOSPI and KOSDAQ indices using LSTM

Introduction | Experiment | **Results** | Conclusion

Comparison Results

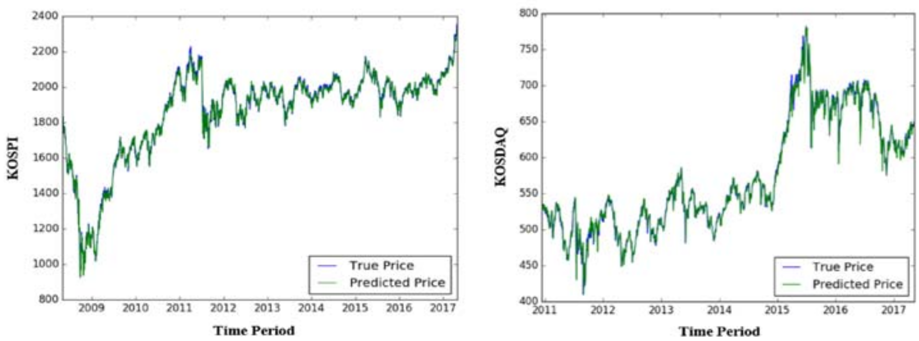


Fig. 5. Comparison predicted price with test period for each Index

Introduction | Experiment | Results | Conclusion

Back-Testing Simulation

Back Testing Strategy

$$P_t = \frac{\tilde{Y}_t - Y_t}{Y_t} * 100$$

1.5<Pt<2: buy 100%
1<Pt<1.5: buy 90%
0.5<Pt<1: buy 80%
0<Pt<0.5: buy 50%
-0.5<Pt<0: sell 50%
-1<Pt<-0.5: sell 80%
-1.5<Pt<-1: sell 90%
-2<Pt<-1.5: sell 100%

in current portfolio capital,

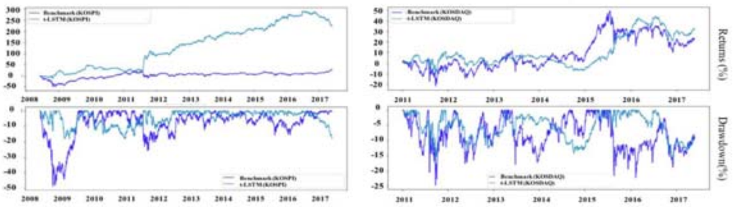


Fig. 6. t-LSTM Simulation results about Returns and Drawdown for KOSPI and KOSDAQ

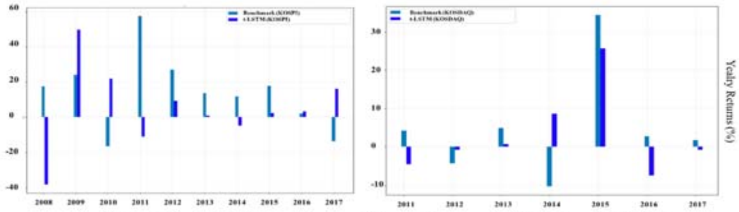


Fig. 7. t-LSTM simulation results about Yearly Returns for KOSPI and KOSDAQ

Introduction | Experiment | Results | Conclusion

Back-Testing Results and Extended Analysis

Table 3 Back-testing Simulation result for each index

	t-LSTM(KOSPI)	Benchmark(KOSPI)	t-LSTM(KOSDAQ)	Benchmark(KODAQ)
Total Return	220.96%	29.33%	32.05%	19.59%
CAGR	14.08%	2.95%	4.54%	2.89%
Volatility	18.99%	14.21%	14.99%	11.8%
Sharp Ratio	99.64%	48.25%	39.35%	48.33%
MDD	-18.25%	-48.81%	-16.86%	-24.77%

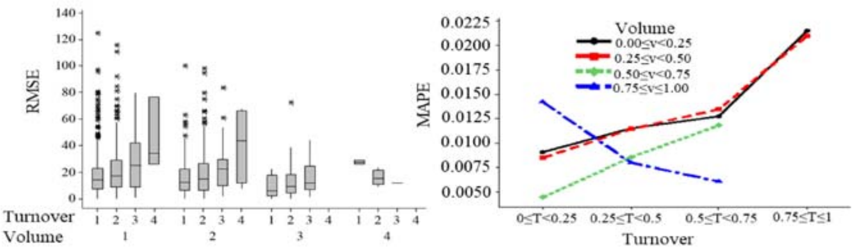


Fig. 8. Prediction error according to turnover and volume.

Introduction | Experiment | Results | Conclusion

Summary & Conclusion

Conclusion

- The presented t-LSTM performs better than simple RNN, BPN or SVR.
- LSTM performs the best (i.e., predicting the closing price using KOSPI and KOSDAQ) when the time step is approximately set between three weeks (15 actual trading days) and four weeks (20 actual trading days).
- The optimal time step of KOSDAQ is expected to be shorter than that of KOSPI, and it implies KOSDAQ is more sensitive to recent information than KOSPI.
- According to the experimental results, KOSDAQ changes at a higher rate than KOSPI as the time step interval changes; therefore, the KOSDAQ has greater volatility than the KOSPI.
- Recall that in both KOSPI and KOSDAQ, the prediction performance worsens as the time step increases beyond the optimal point. Hence, it is imperative to set a more appropriate time step value, rather than setting an arbitrary value as the length of the memory cell in LSTM.
- The performance of t-LSTM with proper adjustment is superior to the benchmark index in returns and risk management.

Thank You