(intel)
Look Inside.™

Using Real-Time Analysis to Build a Financial Sentiment Index with the Intel® Xeon® Processor E7 v2 Family

NTT DATA creates a high-speed, big data–based analysis of the relationship between stock prices and social media sentiments with the Intel® Xeon® processor E7 v2 family

Challenges

- High-speed identification of positive and negative tweets in 35 months of social media data containing tens of billions of tweets
- Build an analysis system for providing a real-time Twitter* sentiment index

Solutions

- Leverage the high-performing Intel® Xeon® processor E7 v2 family to develop a big data analytic solution
- Develop a Twitter sentiment index for financial markets

Impact

- Validated the usefulness and value of a social media sentiment index for financial services companies using historical social media data, and completed sentiment analysis in one day
- Validated the ability to provide a real-time, distributed social media analytic system with the Intel Xeon processor E7 v2 family

NTT DATA

"As hardware advances, analyses that were unthinkable 10 years ago come into reach. We intend to continue working with Intel to supply services with high added value to our customers."

– Keiichirō Nakagawa,
Director,
NTT DATA Mathematical Systems, Inc.

Financial Sentiment Analysis of a Complete Twitter Data Set

NTT DATA Corporation, a member of NTT Group, operates a system solutions business, while its affiliate NTT DATA Mathematical Systems Inc. is involved in consulting, as well as business analytics package development and contracting. The two companies have jointly developed a Twitter sentiment index solution for financial markets. This solution is comprised of systems based on the Intel Xeon processor E7 v2 family. It extracts stock-related tweets from Twitter data and assesses marketing strength or weakness. The system demonstrates the usefulness of social media sentiments in actual market trading.

Information delivery services that use Twitter or other social network service data are operating in the United States, where use of big data for financial analysis is widespread— but there is also growing demand in Japan for such services. NTT DATA signed a contract with Twitter, Inc. in September 2012 for the acquisition and resale of all Twitter data tweeted in Japanese or within Japan. Customers for this data include marketing companies and the news media. "When considering new information services using Twitter data, we directed our attention at financial markets where interest is high, and so we decided to develop this index,"

explains Keiichirō Nakagawa, a director at NTT DATA Mathematical Systems.

Calculated from all the Twitter data in Japan, the Twitter sentiment index is a numerical indicator of positive or negative sentiments expressed in tweets relating to the stock market.

The first step in analyzing the Twitter data is to scan the complete data set, consisting of several tens of billions of tweets, for those tweets that contain stock-related terms such as "Nikkei" or company names such as "NTT DATA." After narrowing down the Twitter data to about one percent of its initial size, the second step brings this total down another two-thirds by identifying those tweets that contain positive or negative terms, such as "The Nikkei is rising" or "It seems too good to be true." The Twitter sentiment index is then defined as positive or negative accordingly. "We needed a faster IT platform for the positivity/negativity identification step in which a text-knowledge extraction technique developed by NTT Group is applied to the large number of tweets," explains Nakagawa.

The companies performed a test by extracting stock-related tweets from 35 months of Twitter data from January 2011 to November 2013, and then analyzing their relationship with stock market indicators. The analysis found a statistically significant correlation

Fast parallel processing capabilities of the Intel® Xeon® processor E7 v2 family help achieve high-speed analysis of a complete Twitter* data set
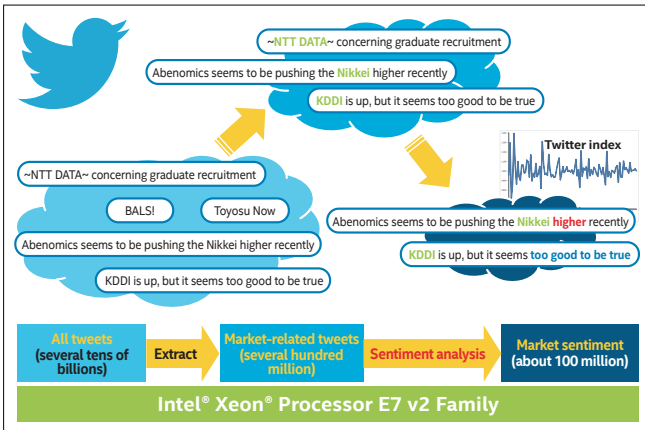
Technique used in sentiment analysis to identify positive and negative tweets

with the Nikkei 225 volatility index (VI), which indicates the views of investors on the likely size of future fluctuations in the Nikkei 225 index. This information could, for example, become part of the decision-making factors in the investment strategies of securities companies, and more efficient investment may be feasible by capturing how the stock market is moving.

Value of Big Data Lies in the Ability to Transcend Barriers Imposed by Time

To achieve high-speed data analysis, NTT DATA configured a test system on a server fitted with four processors from the Intel Xeon processor E7 v2 family, 256 GB memory, and 64-bit CentOS* 6.4. As Koichi Nabetani, senior researcher at NTT DATA Mathematical Systems, explains, "For full-scale testing, we looked to the high parallel processing performance of the Intel Xeon processor E7 v2 family." Masatoshi Yokogawa, section manager at NTT DATA Corporation, says, "We decided to use the Intel Xeon processor E7 v2 family with its large number of cores and high degree of integration with a view to the future provision of the index in real time."

The testing succeeded in performing the positivity/negativity assessment and sentiment analysis of 35 months of data in approximately one day. Nabetani comments, "Along with the improvement in quality that results from being able to complete each analysis quickly, which allowed for more work through the plan–do–check–act (PDCA) continuous improvement cycle, the work also made it clear that our ability to complete our planned schedule of testing several dozen analysis patterns in just one month was in a large part due to the performance of the Intel Xeon processor E7 v2 family." Yokogawa adds, "Our unit performance results showed that a server fitted with a single processor from the Intel Xeon processor E7 v2 family can analyze 3,700 tweets per second. This result shows the potential for real-time analysis. As the real-time processing capabilities improved, it was clear that the server could be used in actual business operations."

Masahiro Kawagishi of NTT DATA Corporation says, "We adopted measures to minimize the total time for parallel processing of large quantities of data while also seeking to balance the load on each CPU core. Making further improvements in parallel processing performance is a challenge for the future." Referring to the business aspects of the testing, Nakagawa emphasizes, "In business, speed is everything. The value of big data lies not just in the large quantities of data but also in transcending the barriers imposed by time—for example, being ahead of other people in learning something that would have taken several days to find out in the past."

Potential Applications Include Assessing Product Reputations and Detecting Complaints

NTT DATA has received many inquiries about the Twitter sentiment index since its release was announced in March 2014, particularly from securities companies. Nabetani comments, "We plan to develop additional Twitter sentiment indices for industries such as retailers or other B2C businesses. We have already started looking into applications outside the financial industry such as reputation analysis and detecting complaints."

As it goes about creating new services and applying advanced analytical techniques to large amounts of data, NTT DATA has high hopes for the Intel Xeon processor E7 v2 family. Nakagawa describes the outlook for the future by saying, "Hardware and analysis algorithms are inseparable technologies for business analytics. As hardware advances, analyses that were unthinkable 10 years ago come into reach. We intend to continue working with Intel to supply services with high added value to our customers."

Find the solution that's right for your organization. Contact your Intel representative, visit Intel's **Business Success Stories for IT Managers**, or explore the **Intel.com IT Center.**

Keiichirō Nakagawa (front left)
Director (Doctor of Engineering)
NTT DATA Mathematical Systems Inc.

Masatoshi Yokogawa (front right)
Section Manager
NTT DATA Corporation

Koichi Nabetani (rear left)
Senior Researcher
NTT DATA Mathematical Systems Inc.

Masahiro Kawagishi (rear right)
NTT DATA Corporation