Using People Analytics to Help Prevent Absences Due to Mental Health Issues

With AI, especially deep learning, experiencing a third boom driven by such factors as the improved performance of computer hardware and falling cost of storage resulting from technical innovation, business is collecting more big data and making greater use of AI to analyze it. This has raised expectations in the field of human resources for the use of people analytics (the analysis of data on employees) in management. Hitachi has responded to these market expectations by combining AI with the knowledge of human resources it has acquired through its Lysithea solution and also the human capital data accumulated by Lysithea. This article describes the commercialization of a solution that uses people analytics for corporate health management.

Tsuneo Morii
Jun Hirayama

1. Introduction

Lysithea, Hitachi's integrated human resources solution that was initially launched as an employment management system, has expanded the scope of its services to encompass talent management, human resources and payroll management. It is currently used by about 1,200 companies with a user base of 1.64 million people. While people analytics (the analysis of data on employees) has attracted interest in recent years, many companies have not yet been able to put it into practice due to the difficulty of collecting data for analysis. Accordingly, Hitachi looked at using the data accumulated by Lysithea for people analytics, launching Lysithea/AI, which integrates artificial intelligence (AI) into the solution, in February 2017.

2. Human Resources Challenges

2. 1 Human Resources Feedback

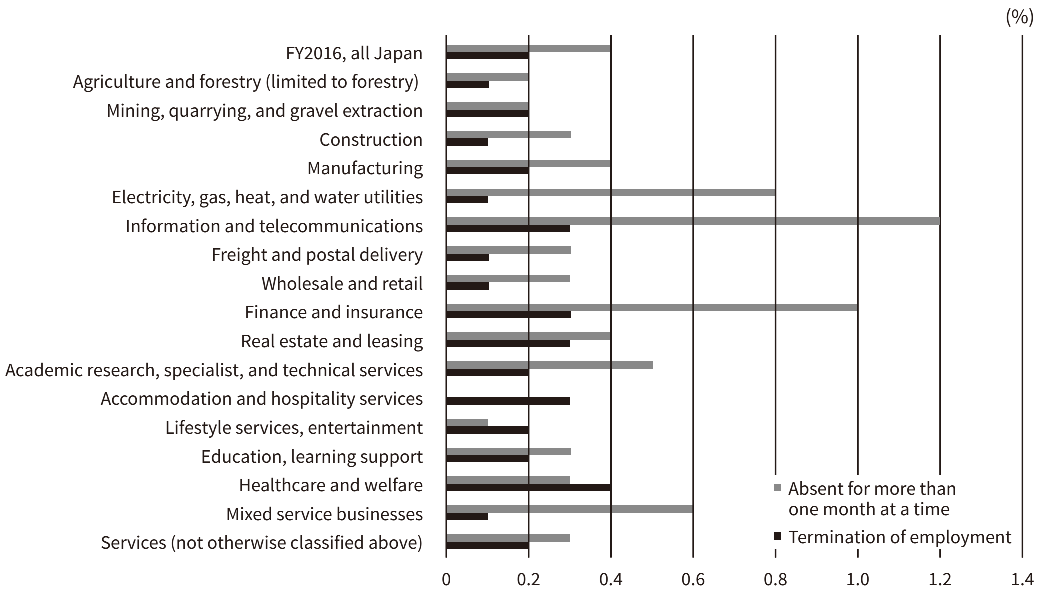
Market research indicates that companies mainly look to people analytics as a way to increase productivity and efficiency. In consultations with Lysithea users, many companies also raise the issue of productivity improvement. Meanwhile more than a few companies talk about people analytics as a practical measure that human resources departments can use to avoid people taking time off due to mental health issues and to prevent young and highly skilled talent from leaving the firm for this reason. In fact, absences due to mental health issues pose a major problem for productivity and efficiency as in many cases the subsequent performance of these people after returning to work remains low or they suffer relapses leading to further time off.

According to the 2016 Industrial Safety and Health Survey (actual condition survey) performed by Ministry of Health, Labour and Welfare, the industry with the highest percentages of employees taking time off due to poor mental health was information and telecommunications (1.2%), following by finance and insurance (1.0%). As the average for the nation as a whole was 0.4%, these two industries stand out (see **Figure 1**). A Cabinet Office report put the cost to companies of each employee taking time off due to poor mental health at 4.22 million yen for each absence.

2. 2
Lack of Prevention

This section uses the example of the safety and health group at Hitachi Solutions, Ltd. to describe what human resources departments can do about the problem of absences due to poor mental health. Whereas the percentage of Hitachi Solutions employees taking time off due to poor mental health had previously been running at about 1%, the company succeeded in bringing this down to 0.77% in FY2016. The greatest effort was put into supporting a return to work. Employees who took time off due to poor mental health sought to obtain treatment and return to work within the time provided by the leave available to them under employment rules. In many cases, however, for reasons that included their own impatience, they returned before fully recovering and experienced relapses within one year. One factor in the 2016 fall in the percentage of people absent for mental health reasons was the effort made to minimize relapses after returning to work, including the use of transition periods during which they could re-establish the rhythms of daily life before resuming full duties. While preventive measures need to be introduced if the percentage of such employees is to be reduced further, there is uncertainty about how this should be done and on what basis, and it is this that is the long-time challenge facing human resources departments. Accordingly, preventing absences due to poor mental health was set as the initial target for the use of people analytics.

Figure 1 — Percentages of Employees Taking Time Off or Terminating Employment Due to Poor Mental Health in Different Industries
The graph shows the percentages of employees who take time off or terminate employment due to poor mental health in different industries.



Source: Summary of Results, 2016 Industrial Safety and Health Survey (actual condition survey), Ministry of Health, Labour and Welfare

3. Use of AI to Predict the Future

3. 1 Data Collection – Data and Causes Assumed by Human Resources

How can people analytics be used as a way to prevent absences due to poor mental health. The first objective was to apply machine learning to a set of benchmark data from employees who had previously taken time off for mental health reasons, and then to use this to predict the probability of future absences for all employees. The choice of variables (feature values) has a large impact on the accuracy of machine learning prediction. Accordingly, factors associated with the onset of mental illness were first collated by drawing on the experience of workplace physicians familiar with the issues surrounding this form of absence. Brainstorming was conducted to identify the factors that lead to people take time off due to poor mental health and the environments and other circumstances that drive the onset of mental illness. **Figure 2** shows the results of this work. Data collection was then undertaken based on these results, with the human resources department and industrial physicians being asked to investigate whether any data existed that could measure these factors and whether this data could be made available.

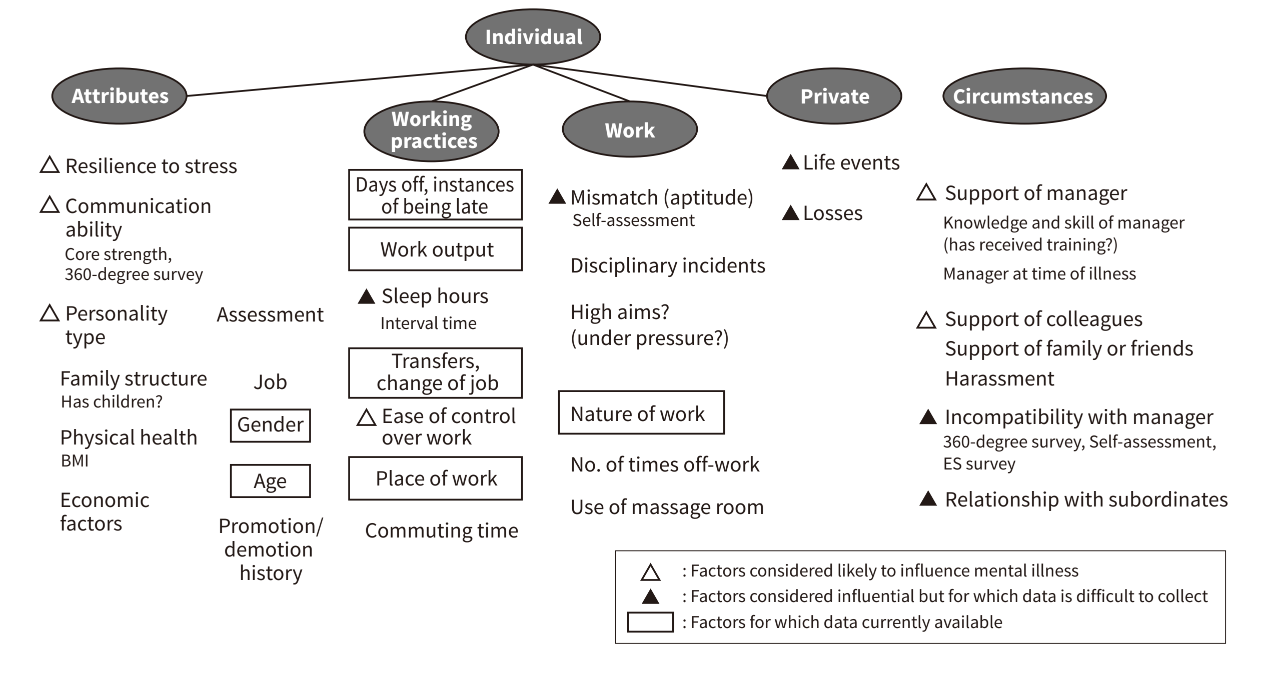
3. 2 Analysis Algorithm

A number of candidate algorithms exist for analyzing the collected data. **Table 1** summarizes these and lists their features.

Although prediction accuracy is important, the aim of the prediction is to identify employees with a high probability of developing mental illness and to provide them with care to prevent this from occurring. However accurate the predictions, they have no value

Figure 2 — Brainstorming Factors that Influence the Onset of Mental Illness

Brainstorming with industrial physicians was conducted to identify the factors that lead to people taking time off due to poor mental health and the environments and other circumstances that drive the onset of mental illness.



BMI: body mass index ES: employee satisfaction

* 360-degree survey: A form of evaluation also known as multi-source feedback. The individual's workplace behavior is observed and assessed by supervisors, subordinates, and colleagues.

Table 1 — Machine Learning Algorithms
These algorithms were considered as candidates for analyzing the collected data.

Algorithm	Description	Predicts probability	Provides reasons
Logistic regression	A technique that uses regression for classification. It is a regression model for binary classification problems that works by using the output of linear regression as an input to the logistic function. Mainly used for classification in statistics, and also frequently used in medicine and social science.	✓	✓
Random forest	A prediction method that generates a large number of decision trees and collates their respective results. The individual decision trees are kept separate and the method learns by sampling from the explanatory variables and from the learning data to give them different characteristics. Used for predicting competitive results.	✓	✓
Gradient boosting	A prediction method that generates a large number of decision trees and collates their respective results. It is the most popular method and works by successively increasing (“boosting”) the number of decision trees, updating the labeling of cases where existing decision trees give incorrect results and using this to generate new decision trees.	✓	✓
Support vector machine	A technique that uses regression for classification. It creates a two-category regression model that maximizes the margin between the two categories to achieve a high level of generalization even for small amounts of data. The downside is a long learning time. It is a type of learning model with excellent recognition performance.	✓	n/a
Neural network	A mathematical model that seeks to use a computer simulation to replicate a number of characteristics seen in brain functions. Neural networks can often produce good solutions with a comparatively small amount of computation to problems involving multi-dimensional data such as images and statistics where linear separation is not possible. Applications include image recognition and the use of analogical inference on purchases based on customer data in markets.	✓	n/a
Deep learning	A machine learning technique that allows computers to learn tasks that people perform naturally. The technique is known for progressively deeper learning of the characteristics embedded in data achieved by using multiple layers of neural networks.	✓	n/a

on their own because it is not possible to take action without also knowing what it is that is putting the person at increased risk. Accordingly, to prevent the prediction process from being a black box that does not indicate the contribution of each factor to the result, algorithms that could provide this information were trialed and the one with the best accuracy was selected. This was the gradient boosting algorithm.

4. Analysis and Prediction

4. 1 Characteristics of People Taking Time off Work – Use of Statistical Analysis to Identify Feature Values

The collected data was first prepared by filling in missing values or excluding people who had missing data and then performing statistical analysis to identify which variables distinguished people who had been absent for mental health reasons from those who had not. This was then converted into abstract data structures, including the relative variations between different variables obtained by normalization, and the feature values were obtained by further reducing and combining variables by dimensionality reduction and dimensionality compression.

4. 2 Increasing Size of Data Set

As the data used to predict who will take time off for mental health reasons also contained information on people’s working environments and other employment details, the analysis obtained feature values specific to the companies they worked for. Accordingly, it is believed that performing analyses on data from each company separately will lead to better prediction accuracy. Unfortunately, the amount of data is an issue when performing machine learning. That is, the smaller a company, the less benchmark data is available to work with. A company with 10,000 employees, for example, will only have about 100 people a year taking time off for mental health reasons. While it is possible to use such data from past years, the current wave of work style reforms are bringing major changes to working conditions, so the use of older data risks degrading performance.

The data used in analysis includes both time-series data such as the number of overtime hours or holiday days taken and one-time data that can only be collected one or two times a year, such as performance evaluations or questionnaires. For time-series data, multiple data ranges for each individual were obtained by using data from different time periods within the overall time period designated for the analysis. By combining this with the one-time data to create multiple data sets, Hitachi Solutions was able to increase the amount of data available.

4. 3
Results and Performance

A prediction model was produced by performing learning on the data obtained by the procedure described above. This prediction model was then applied to data that was not used in developing the model to assess how well it could identify employees who had, in fact, been absent for mental health reasons. The following gives an example of how this was done.

This example used time-series data covering six months. The data for each individual was shifted back in time in one-month steps to create data equivalent to six people. Similarly, the stress check responses used in the analysis were converted to simple numeric values, such as "response to question one = 3," without including the question meaning. When used on data for the time period of interest, the prediction model identified 134 people as having an elevated risk of absence due to poor mental health, of whom 43 had actually taken time off for this reason. Although the output of the prediction model is the probability that an individual will take time off in a given month, a prediction was judged as being accurate if the person took such time off within the following year.

The mean accuracy of correct predictions was about 30%. Given that the percentage of employees actually taking time off due to poor mental health is only around 1%, the fact that about 30 of the top 100 or so people who were identified as high risk by the prediction turned out to be people who actually had taken time off was deemed to be a good result and sufficiently accurate for practical use in offering early intervention.

5. Lysithea/AI – Stress Prediction Service for Organizations

5. 1
Stress Prediction Service for Organizations

Hitachi Solutions has commercialized Lysithea/AI as a solution for using the data manipulation and machine learning described in the previous section to predict which employees are similar to those who have taken time off in the past due to poor mental health (see **Figure 3**). The solution consists of two services. The first service analyzes customer data to produce a prediction model, assesses the prediction model's prediction accuracy, and also offers consultation to show customers where best to direct efforts by identifying the characteristics of employees who have been absent due to poor mental health (the factors and conditions that strongly influence mental illness). The other is an ongoing service that facilitates early intervention by using the prediction model to identify, on a month-by-month basis, which employees are likely to take time off work based on the latest data.

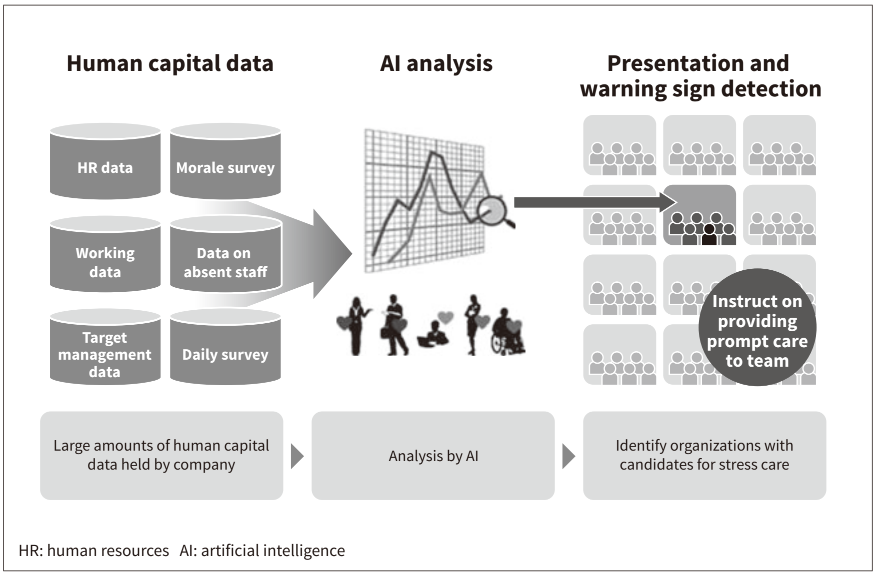
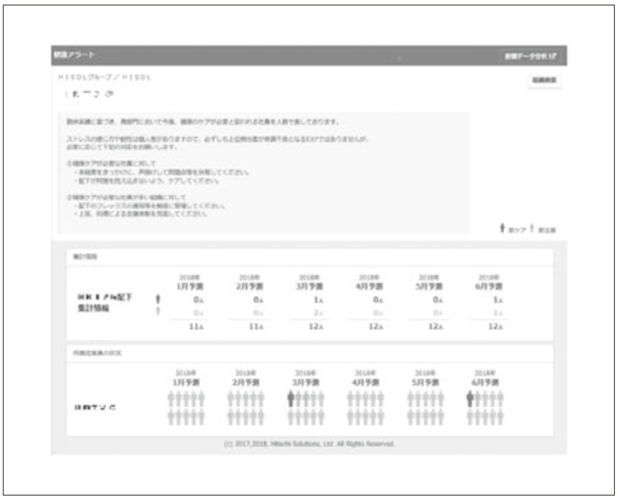


Figure 3 — Overview of Lysithea/AI
The figure shows an overview of how the Lysithea/AI solution works.

Figure 4 — Example Lysithea/AI Screen

The screen shows the analysis results in terms of health alerts and working practice rankings based on the predictions.



To improve accuracy, the ongoing service has systematized the process of using the latest collected data to revise the model each month based on current working practices and other aspects of the workplace environment and to issue new predictions. This allows the service to make predictions each month based on the latest data about which employees are likely to take time off in the following two months.

5. 2
Privacy Considerations

Section 4.3 referred to the use of stress check response values, something that Hitachi Solutions analyzes and utilizes at the request of industrial physicians. Use of this data is limited to what these physicians are permitted to use. In general, employees must give their consent before data can be used in analysis. However, even when consent has been given for the analysis, because the analysis works by identifying people at high risk of absence due to poor mental health, and given that this has the potential to influence how the people concerned are treated and assessed, appropriate consideration needs to be given to the question of who should be given access to the results, and the individual's consent to this access should be obtained.

The ongoing service provided by Lysithea/AI includes a facility for controlling who gets to see information based on the customer's circumstances.

5. 3
Presentation of Results and Triggers for Action

Lysithea/AI facilitates early intervention for employees at high risk of absence due to poor mental health. However, as the manager directly involved in the employee's work is in the best position to offer this care, such interventions are unlikely to occur unless the manager is informed of the predictions. Hitachi Solutions launched Lysithea/AI in October 2017 and the current product includes this as a standard feature (see **Figure 4**).

The characteristics of employees who take time off due to poor mental health identified by data analysis include the number of days of paid leave they take, how frequently they are late (either late to arrive or late to leave work), and the number of times they work on a public holiday. As these factors are seen as strongly influencing mental illness, the service works by making the respective variables visible and informing managers of when an employee's lifestyle rhythms are disrupted so that action can be taken. As employees who take time off tend to take a large number of days of paid leave, the service displays a ranked list of people in the organization who have taken a lot of leave days to provide greater opportunities for managers to talk to these people and ask about things like their state of health or personal circumstances.

6. Conclusions

Many different factors are involved in the onset of mental illness and there is a limit to how much data can be collected in practice. In the above example, the analysis did not have access to data about individuals (such as their resilience to stress or personality types) or other private data (such as losses or other life events). As the ability to collect meaningful variables is essential to improving the accuracy of machine learning, greater accuracy should be possible through the routine collection and use of such information as aptitude test results, survey responses, and sensor data, for example. However, recognizing that advances in analytics on their own will be insufficient to resolve the question of which preventive measures can be adopted to reduce absences due to poor mental health, especially given the sensitivity of these data for privacy reasons, this is something that will require collaboration between human resources departments, line managers, and industrial physicians, and the consideration of new practices based on understanding and trust between management and employees.

References

1) Gender Equality Bureau, Cabinet Office, "Benefits Companies Gain from the Balance between Work and Life," Council for Gender Equality, Council of Executives of Public and Private Sectors to Promote Work-life Balance (Apr. 2008) in Japanese.
2) Ministry of Health, Labour and Welfare "Overview of Survey on Industrial Safety and Health Survey (Actual Condition Survey) in 2016," Japanese Government Statistics (Sep. 2017) in Japanese.
3) PwC Consulting LLC, ProFuture Inc., "Report of People Analytics Survey 2017," (Nov. 2017) in Japanese.

Authors

Tsuneo Morii
HR Technology Center, Lifestyle Innovation Division, Hitachi Solutions, Ltd. *Current work and research:* Review and trialing of new services (such as employee attrition prevention support, placement matching, etc.) following after "Lysithea/AI," helping prevent absences due to mental health issues.

Jun Hirayama
HR Technology Center, Lifestyle Innovation Division, Hitachi Solutions, Ltd. *Current work and research:* Review and development of dashboards for effective visualization and operational support for "Lysithea/AI" results, etc.