CASE STUDY

Improving materials data quality

Executive summary

Selecting materials with the right properties for a project is critical. Matmatch makes this process easier by providing comprehensive databases of materials and their properties, with data sourced from materials suppliers. In practice, however, not all characteristics are known for each material, which is why Intellegens' machine learning tools can add value. Alchemite™ can help to fill the gaps, detect outliers, support material selection, and even identify the gaps in the property space that might be targets for new materials development.

This case study describes the ability of Alchemite™ to **populate missing material property data** and add significant value to Matmatch services. It also shows that the Intellegens machine learning tools were easily integrated with Matmatch's systems and that their data scientists could **effortlessly build complex models of multi-dimensional data** with Alchemite™ software.

Intellegens Ltd., Eagle Labs, Chesterton Road, Cambridge, CB4 3AZ, UK

Challenge

Material design is a highly complex endeavour due to the vast amount of possible compositions and process parameters and to the time and cost associated with experimentally measuring material properties of interest. Selecting a material that achieves the required set of properties (e.g., strength, corrosion resistance, weight, or cost) is essential for the success of a project.

Selecting a material that achieves the required set of properties is essential for the success of a project.

To help select optimal materials, Matmatch provides extensive databases where characterised materials can be browsed, compared, and matched to customer project needs. However, these databases are sparse, as not every material property is available for each material. The goal of this project was to use the Alchemite™ machine learning toolkit to **impute (populate) the missing data** from the databases as well as **to verify the existing data by identifying outliers**.

Solution

Alchemite™ is Intellegens' unique deep learning tool, which uses the power of deep neural networks to build comprehensive models from sparse and noisy data. Ceramics, polymers, aluminium, and metals datasets were used to test Alchemite™'s ability to create accurate models despite there being over 90% of missing target data.

Four important use cases were identified where Alchemite™ added significant value to Matmatch's databases:

Filling the gaps

Imputation by Alchemite™ can fill in the missing data not provided by material suppliers. This can be presented to clients to help them understand material performance.

Search engine optimisation

Materials may be missing from search results because the property of interest has not been measured. Users would benefit from having potentially suitable, but previously excluded, materials suggested if the predicted property values meet their requirements.

Material supplier suggestions

Clients could search for properties not measured in material supplier datasets, using Alchemite™ predictions instead. This information can be communicated to the supplier allowing commonly required properties to be measured and added.

Requirement for new material design

Matmatch may identify certain search patterns that require the design of new materials. Alchemite™ can help guide this process.

Outcome

Intellegens trained a model on a dataset with metals information. This was a large dataset with 55,000 rows, 254 inputs, and 162 outputs, but more importantly, 96% of the target data was missing. Despite the lack of data, Alchemite™ produced a model in which the coefficient of determination against key targets (defined by Matmatch) was 0.82, which is indicative of very high accuracy.

Additionally, Matmatch data scientists used Alchemite™ to train models on polymer and ceramic databases. Not only did they find that the machine learning tools provided by Intellegens were easy to use and integrate with their systems, but these tools also outperformed their expectations. The aim was for 10-12 material properties to be reliably predicted and, actually, accurate predictions were achieved for twice as many properties.

"We were most impressed by the results we were able to achieve using the Alchemite tool to improve the Matmatch material data set. The outcome exceeded our expectations and we succeeded in predicting significantly more properties than expected across different material categories despite starting with a very sparse dataset."

- Melissa Albeck, CEO at Matmatch

Key conclusions

- Alchemite™ was able to populate sparse and missing data and model complex multi-dimensional data.
- Data scientists from Matmatch found it easy to access the full power of Alchemite™ via the Alchemite™ Engine API, which integrated well with their systems.
- Alchemite™ provided a significant value and potential for additional revenue to Matmatch through data checking, estimation, and gap filling.

Future opportunities

This case study demonstrates that Alchemite™ machine learning can be used to revolutionize the way that material databases are populated and verified. It also establishes the system's potential as a solid platform to maintain and continually improve any database based on experimental information.

Three future opportunities offered by machine learning are:

- Outliers can be detected by the Alchemite™ models, reducing the impact of measurement errors.
- As new data becomes available the Alchemite™ model and its predictions can be retroactively updated and improved.
- Alchemite™ can be integrated with Matmatch's systems in order to automatically populate and curate new databases.

About Matmatch

Matmatch is a materials search platform that connects engineers and material suppliers through the most comprehensive materials database in the world. They want to inspire people to build better products by changing the way the world discovers and uses materials.

About Intellegens

Intellegens has developed a unique deep learning engine, Alchemite™ for training neural networks from the sparse and noisy data typical of real-world science and business challenges. The technique was first developed at the University of Cambridge where it has been used to develop aerospace alloys, guide the design of new drugs, and design next-generation battery technology. The tool is now being used to solve a wide range of industrial customer problems, optimising products and processes, saving time and cost in discovery and development, and enabling breakthrough insights.

www.intellegens.ai | info@intellegens.ai | @intellegensai

