Application Of Case Based Reasoning Technique For Advice On Quick Tap Practice In Tata Steel

Mariana Adderley,

Principal Researcher,
Tata Steel RD&T,
Moorgate, Rotherham, S60 3AR, UK

Mariana.adderley@tatasteel.com

Abstract Modelling complex behaviour of steelmaking processes, affected by the dynamic environment and subjective actions of human interaction is a challenging task. Using modelling techniques, based on first principle, to describe the dynamics of the process, comes with inherent limitations when trying to fit experience into equations. The UK based branch of TataSteel RD&T has successfully developed an application of Case Based Reasoning technology for predictive modelling in the steel industry. The model predicts final chemical composition and advises on safe tapping operations. This paper describes the approach and structure of the model and share experience gained during model development and exploitation. The model evaluation is based on 10 years of continuous monitoring and shows consistent accuracy of 80%-90% under changing manufacturing conditions and material quality. The benefits of using recent experience for effective tracking of changes in manufacturing practices, material quality and qualitative fluctuations are discussed.

Keywords:

Steelmaking process, Predictive modelling, Case Based Reasoning, Knowledge Base

1 Introduction

Basic Oxygen Steelmaking (BOS), also known as Linz-Donawitz-Verfahren steelmaking or the oxygen converter process is a method of primary steelmaking in which carbon-rich molten pig iron is made into steel. The process is a batch process and begins with charging the steelmaking converter with scrap metal, lime based fluxes and hot metal from the blast furnace. The total charge weight for a batch is 180 – 350t depending on the converter size. Then water cooled oxygen lance blows 99% pure oxygen with supersonic speed onto the steel and iron, igniting the carbon, dissolved in the steel and burning it to form carbon monoxide and carbon dioxide, causing the temperature to rise to about 1700°C. This melts the scrap, lowers the carbon content of the molten iron and helps remove unwanted chemical elements. Near the end of the blowing cycle, which takes about 20 minutes, the temperature is measured and samples are taken. The samples are tested and a computer analysis of the steel given within six minutes. The BOS converter is then tilted and the steel is poured into a giant ladle. This process is called tapping the steel. One of the objectives of the BOS batch process is to tap steel with required chemistry. The model presented here predicts the final chemistry of the steel and advises on successful tap i.e. within the target composition. This allows the control operators to save up to six minutes per batch by tapping the steel before the computer analysis arrives. This operation is known as quick tap and it is very important practice within the steel industry when the objective is to maximize the production volume. Without a model, the practice relies on the experience and intuition of the operators.

This paper describes the architecture and operational capabilities of the model, applied to assist the Quick Tapping practice. Analysis of the model results has revealed excellent robustness and consistent level of accuracy over long period.

Beginning with Section two, there is a brief overview of some AI techniques, commonly used for modeling steelmaking processes and outline of the challenges in modern predictive modeling. Section three describes the case based approach to Quick Tap modeling and gives details of the model's architecture. Section four presents the evaluation and results of model exploitation. The final Section five closes the paper with conclusions and directions for future work.

2 Predictive modelling and challenges

In early 1998 there was a need in the company to build a robust model that estimates the final chemical composition from a complex reactor and advises on hitting the target levels. The model was to produce a prediction at all times and respond adequately on changes.

Existing research on application of AI techniques for predictive modelling in steelmaking at the time has been focused particularly on using Artificial Neural Networks (ANN). In addition, the use of modern regression and optimisation techniques has been reported [1]. One problem that was not addressed with the latter was the issue of the inherent uncertainty in the data and the effect of non-measured changes in the working environment, human behaviour or raw materials quality. It was the special features of learning and generalisation that made ANN one of the most accurate predictive methods. However, a difficult problem with learning in many applications was that the predicted elements may depend on some hidden context, not given explicitly in the form of predictive features [2]. Often the cause of change is hidden, not known in advance, making the learning task more complicated. Changes occur over time. For example, changes in raw materials quality can have a big effect on the quality of the final product. Often the old experience can become irrelevant to the current time and thus the learned knowledge can be out of date. Changes in the hidden context can induce changes in the target concept, phenomena generally known as 'concept-drift' [2]. An effective learner should be able to track these changes and adapt quickly to them. Even if the target concept remains the same but the data distribution changes, a model rebuild may be necessary as the model error may no longer be acceptable.

The most common concept-drift handling technique was based on instance selection and involved generalizing from a window that moves over recently seen examples. It used the learnt concepts for prediction only in the immediate future. A version of this technique was employed in the model.

3 CBR approach

Data analysis from two year historical examples revealed that clusters of batches with similar behaviour before tap were closely related to clusters of batches with similar deviation from the aim tap analysis after tap. In other words, it was possible to map experience before with experience after tap [3]. This led to the idea of experimenting with CBR model.

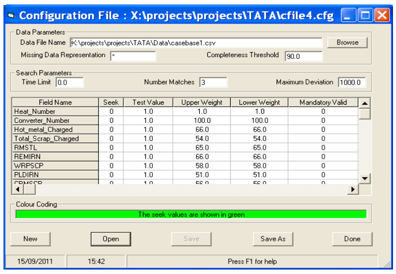
CBR technology was chosen after fast prototyping feasibility study. The study involved ANN, Fuzzy rule-based technology and CBR and concluded that a hybrid model with CBR as a core technology was the best suited modelling technique for this particular application. Table 1 shows the accuracies achieved with each technique on unseen data from the same time period as the training data. These accuracies are defined as the percent correct predictions for a binary 'yes' / 'no' on target advice.

Table 1. Model accuracies for experimented techniques

Technique	Percent accuracy
First principle	86%
Rule based	87%
ANN	96%
CBR	90%

The similarities between CBR and ANN in their generalisation capabilities were used to achieve the required accuracy. The attributes of the CBR system allowed flexibility and assured low maintenance of the model. Finally, continuously learning from the constantly changing environment, accounted for any variations of material quality, production practices and unknown factors.

The approach was to use the in-house developed Nearest Neighbour tool [8] for retrieval and a rule based module for adaption of the proposed solution. The Nearest Neighbour tool consisted of two modules: Query Agent – a tabular editor (see Fig.1) user interface and Rolling Neighbour – a modified version of the k-Nearest Neighbour algorithm using a 'rolling' casebase. Fig.2 shows one of the output displays in manual mode of the tool.



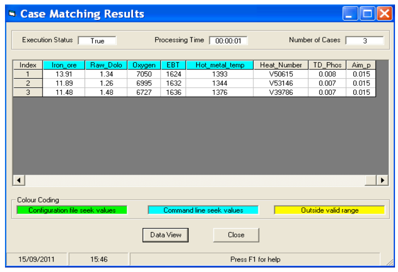


Fig. 1. Query Agent editor

Fig. 2. Rolling Neighbour output display

Each batch was a case, representing experience before tap (case description) and actual tap analysis (solution). Using the actual chemical analysis of the retrieved batches, the target case chemistry was estimated and the probability of success to be within the acceptable limits was produced.

The model functionality is shown on Fig.3. When a description of a new case (batch) is passed to the model, a search for similar cases in the case base begins. The most similar 'k' cases, matching the current description are retrieved. Case "Retrieval" is the first stage of the CBR cycle and is crucial for the successful operation of any CBR system. The retrieved cases present several possible solutions within certain range, therefore the "Reuse" strategy involves adaptation procedures using statistical estimation of the tap analysis and calculation of probability for successful quick tap, corrected with rules from a knowledge base.

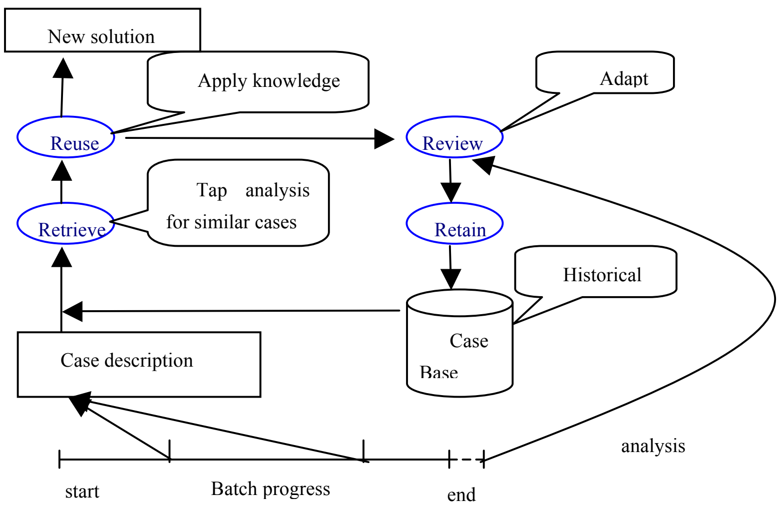


Fig. 3. CBR cycle applied to the Quick Tap Predictions task [4]

Subsequent stages include verification of the proposed solution (i.e. review how close the actual value was to the proposed prediction and the correctness of the advice) and "Retention" of the new experience for future reuse. The "Review" stage of the model assures continuous performance monitoring and adaptation. The "Retain" stage guaranteed continuous learning from experience.

3.1 Case representation

The case is a collection of experience, represented as description of a situation and solution. Applied to the Quick Tap model domain the case represented the batch history as a collection of information and measured parameters and had the structure, shown in Fig.4.

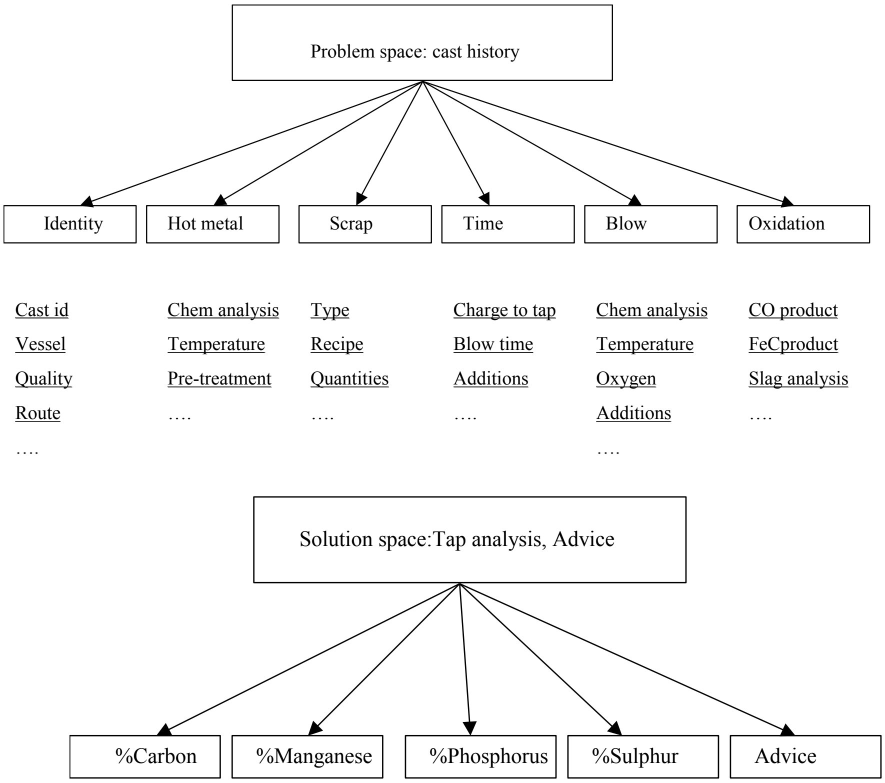


Fig. 4. Quick Tap case representation

The problem space was divided between six groups, describing batch behaviour from a different perspective. Grouping was done in the way a domain expert will look into the problem in order to explain a particular situation. Each group was split into subgroups. This grouping of inputs was important, when the solution space expanded to include four elements, because it allowed task decomposition and enabled retrieval of each sub-group separately. The main advantage of this approach was that it enabled the CBR system to use different parts of cases in the case base to solve different subtasks (groups of the solution space) for the new problem. The model allowed each batch to have different number of parameters, describing the problem state, depending on how many parameters were available at the triggering point. Prediction was triggered twice during the process at two different time intervals before the end of the process (see Fig.3). The first (early) run of the model reliably predicts 20% to 25% of the batches for successful quick tap and gives early warning for corrective action, i.e. the model at this point was very conservative. The second run of the model produces predictions before the final composition analysis is available. At this stage 90% of the batches have the correct advice. The wrong advice (False Positives) is around 5% and missed opportunities (False Negatives) around 5%.

The accuracy of model predictions can be seen on Fig.7 for the 10 year period 2001-2011. The ‘similarity measure’ represents the percentage of correct predictions.

3.2 Feature selection

The features for each case were identified initially using the whole history of the batch from the beginning to the end of the batch. This resulted in a large number of features and the danger that some of them may distort the results. Therefore feature selection was necessary to reduce the dimensionality of the feature space. Information Gain (IG) and Sensitivity Analysis (SA) were used to select the most predictive features. The cross validation experiments, varying between fifty-two and four features across nine data sets, indicated that the optimal number of features is plant specific and varies between eight and fifteen aiming for minimum accuracy of 80% correct predictions. Use of a minimum number of features without sacrificing the quality of the final result also reinforced the robustness of the model.

3.3 Case retrieval

In the Quick Tap model a pre-defined indexing was used, that acted as a filter, enabling the CBR system to locate the relevant subset of cases. The most similar ‘k’ cases to the current problem were then identified from this subset by applying k Nearest Neighbour Algorithm (k-NN).

A modified version of the k-NN similarity retrieval algorithm was used for greater flexibility and computational efficiency. Depending on information available at the time of prediction, the target case can have different set of features used to find a match. To represent the minimum set of features required to produce a reliable result the attribute ‘mandatory set of features' was introduced. The 'maximum deviation' model parameter allowed setting the threshold after which the cases are not considered similar. Similarity was calculated for each feature individually as a deviation from the target case. The case similarity was a weighted sum of individual feature similarities, where the weight reflected the importance of the feature and relevancy of the case. In this context the similarity measure reflected the deviation from the target case, therefore the case with the smallest value for similarity (deviation) was considered as the best match.

Probabilistic theory, combined with pre-defined rules from a Knowledge Base (KB) was used to generate the summarisation of the final solution in a form of an advice: "Wait for sample", "OK to tap" or "Caution ". The KB was initially created with several basic rules that an expert will use in order to make a decision on successful Quick Tap. Gradually with the exploitation of the model the KB was expanded. The extra rules added were plant specific.

3.4 Fine tuning

A serious problem for the Quick Tap prediction model was the occurrence of False Positives (FP), where the predicted value was incorrectly classified as within the target range for quick tapping. A FP is significantly more serious than False Negative (FN), where the predicted value is incorrectly classified as outside the target range. The occurrence of FP was unacceptable and due to this fact, the model had to be fine tuned to the optimal number of retrieved cases. To maximise the accuracy of the retrieval algorithm it was acknowledged that when two values are compared sometimes the effect is different when the first value is bigger or smaller than the second value. For example, to compare the value of 0.6% hot metal Silicon with 0.55% and 0.65%, in terms of final Phosphorus. The value of 0.65% is 'closer' to 0.6% than 0.55%. This is determined by the behaviour of the cast with hot metal Silicon above and below the target value i.e. the pair of casts with hot metal Silicon levels 0.6% and 0.65% will behave more similarly in terms of final Phosphorus level than the pair with 0.6% and 0.55% Silicon. To make this distinction, different weights were set, depending on the sign of the difference between target and compared value. This method was used to fine tune the model and optimise the FPs and FNs.

3.5 Case-base Management

When the final tap analysis for a case was available, all known information for this case was assessed for completeness and added to the case base as a new case. The oldest case is removed from the case base when a new case is added. If the new case had FP or FN prediction, the retrieved cases used for this prediction are assessed and the case with the biggest contribution for this misclassification is marked as unsuitable neighbour for the new case. If the same case is identified as 'unsuitable neighbour' on five occasions, then this case is deleted from the case base as ‘noise’. The 'noise' casts are kept in a log file for further off-line investigations.

During the development stage a trial exercise was completed to compare a ‘rolling’ case base with

(a) ‘growing’ case base that contained every new example;

(b) ‘interesting’ case base that assessed every new case and assign it to a cluster of cases in the case base. If the case belonged to a big cluster and was very similar to a case already in the case base (and with the same outcome), it was not considered as ‘interesting’ and therefore not added to the case base.

The comparison showed that the ‘rolling’ case base gave consistently better results.

4 Evaluation

The key objective of this evaluation is to observe the performance level over a long period of time and to assess how the model handles data uncertainty, missing data and concept drift.

The evaluation data presented here, are selected from ten year period of predicting Phosphorus level only (the model was originally developed for predicting Phosphorus). The predictions are from the second trigger (see section 3.1) of the model. To observe the model performance parts of datasets from 2001 to 2011 are shown. Dataset1 is a collection of data from January 2002 to December 2002 and represents stable performance of the model two years after implementation, when there were no major disturbances within the process. Dataset 2, 3 and 4 are a selection of data from May 2004 to May 2011. This represents a period of time of several of changes, concerning mainly distributions of inputs and material quality. It is interesting to see how the model handles these changes and gradually learns from collected new examples. The case base at any point of time contains cases from one year backwards.

For a quick visual assessment of the model performance the plot of cumulative frequencies of the error in chemistry estimation (see Fig.5) was used, but this only accounts for part of the solution.

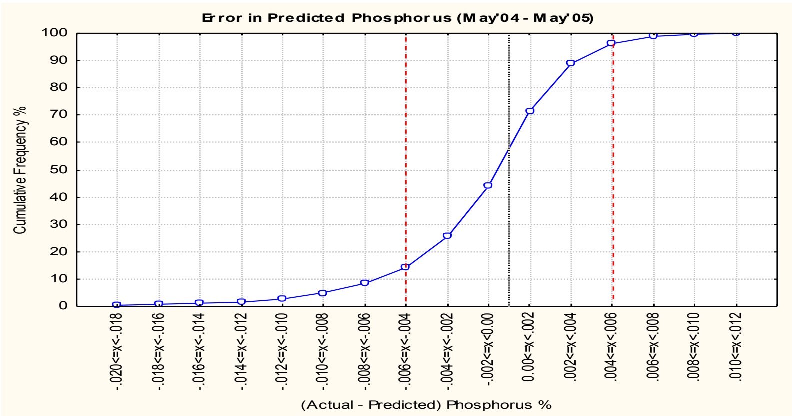


Fig. 5. Cumulative frequency of Phosphorus estimation error

In order to assess the correctness of the advice a complex measure of similarity between actual and predicted solution was used.

Consider the examples in Table 2, listing the difference between actual and model solutions for Phosphorus only.

Table 2. Comparing model and actual solutions

Aim %P	Actual %P	Predicted%P	Advice	Probability To hit target	Verdict
0.02	0.017	0.0182	OK to Tap	100%	Correct(TP)
0.012	0.011	0.0141	Wait sample	25%	Missed(FN)
0.012	0.013	0.0137	Wait sample	30%	Correct(TN)
0.02	0.023	0.0173	OK to Tap	85%	Wrong(FP)

The first case has actual Phosphorus level below the aim and the model advice is correct for safe quick tap (True Positive /TP classification). The second case could have been successfully quick tapped, but the advice was against it. This represents the missed opportunities or FN classification. The third case was correctly advised against quick tap i.e. True Negative (TN). The last case was wrongly advised to quick tap i.e. FP classification. The objective of the fine-tuning of the model was to minimise the FP and FN, but specific attention was paid on minimising the FP as they are unacceptable practice.

Many CBR developers use the overlap similarity metric (Sim) as a measure of the system accuracy. Here Sim is the number of batches, correctly advised, divided by the total number of batches:

$$Sim = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Figure 7 shows the level of similarity measure over ten year period. It is in the region of 0.8–0.9 and with a couple of exceptions hardly drops below this level. This demonstrates the stability of the model over the given time period.

From production effectiveness point of view the aim is to reduce the TNs and FPs. Therefore, if the plant is operating effectively the TNs and the FPs will be negligible in comparison to TPs. With this in mind traditional Information Retrieval (IR) measures to report performance [6] were employed.

Precision (P) is the proportion of proposed batches for quick tap that are actually correct

$$P = (TP) / (TP + FP) \quad (2)$$

And

Recall (R) is the proportion of actual target hits that are correctly advised for quick tap

$$R = (TP) / (TP + FN) \quad (3)$$

Generally high Precision is desirable because lower FP and higher TP is important production quality indicator. The combined effect of these two measures is commonly known as the F-measure (F) calculated as their harmonic mean

$$F = 2PR / (P + R) \quad (4)$$

By applying this measure, far more realistic comparison between actual and predicted solution was achieved. The graph in Fig.6 illustrates the ability of the model to handle small (Dataset1) and large (Dataset2) unseen disturbances to the problem (input) space. Dataset1 represents a stable period of operations, during which not many disturbances were recorded. In the case of Dataset2, in the second half of 2004 a series of changes were introduced e.g. new equipment, production practices, and product changes and is evident that the performance indicators of the model deteriorated (increase of FP) for a while until the model learned from the new experience and restored the stable performance. It has to be noted that after the stable period of 2002 there was no external intervention to the model, i.e. the model operated completely autonomously. Dataset3 represents period of stable operation, again with levels of accuracy similar to Dataset 1. Dataset 4 is interesting, because it represents a period of time where there was no regular feedback, i.e. no regular update of the casebase. This affected the accurate representation of recent experience and therefore affected the accuracy of the model prediction (increase of FP). During this period the model predictions are not based on recent experience and when compared with Dataset3 the benefit of using it is obvious. It is also interesting to see that the level of Recall and respectively the number of FN have not been affected greatly by disturbances in the input space.

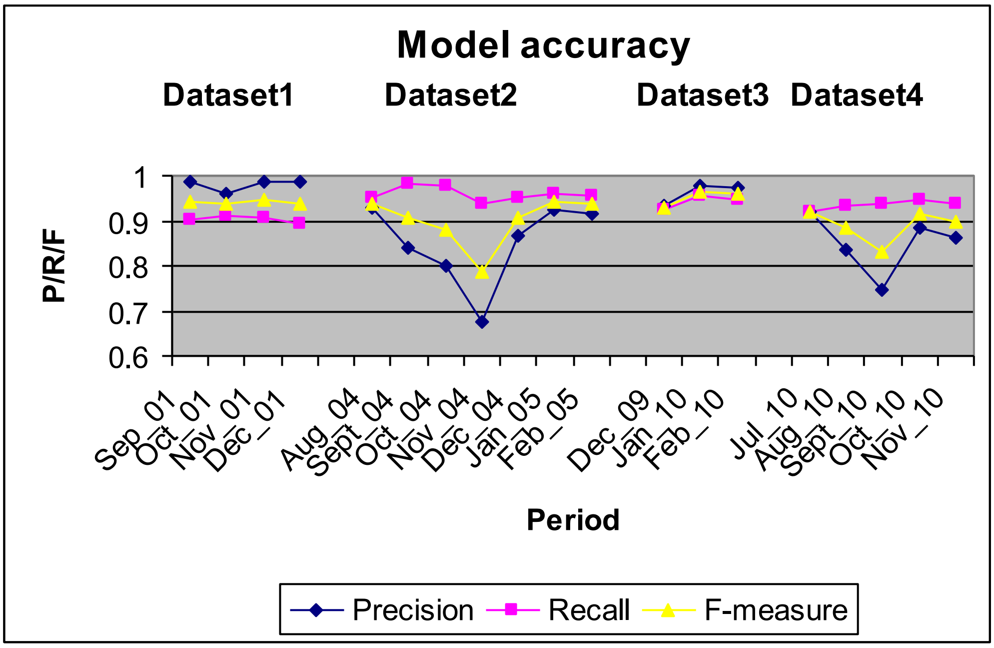


Fig. 6. Model accuracy 2001-2010

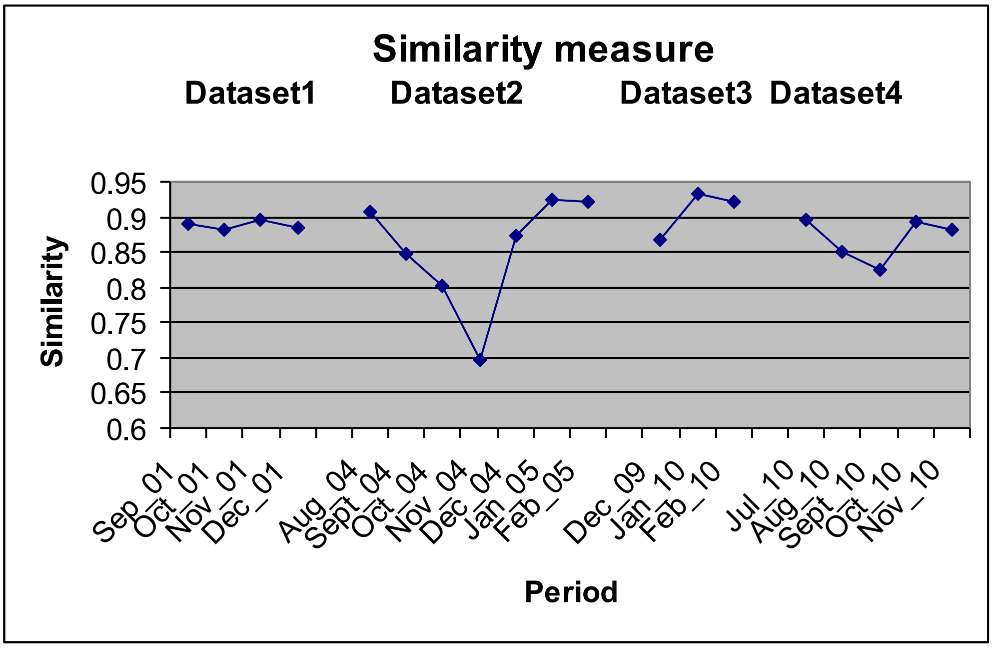


Fig. 7. Similarity measure for a period of Sep 2001 – Nov 2010

5 Conclusions

Ten years exploitation of the Quick Tap model has proved that the CBR technology can be successfully applied in predictive modelling. The experience in using this technology is summarised below:

- Short development cycle: there was no need to understand fully the problem, only to collect examples, which avoided the common bottleneck of knowledge elicitation. This was true during the initial development. For the fine tuning of the model, good case representation and knowledge of the problem space was crucial.
- It was very important to include the misclassifications of the model in the learning process in order to keep consistent accuracy level;
- Small number of mandatory inputs gave better robustness;
- The concept of learning from experience was easy to understand;
- Decision, based on experience can account even for factors that can't be measured;
- Model, requiring low maintenance (virtually no maintenance) and delivering high level of accuracy very quickly earned the respect of the users;
- The model gained the biggest approval during times of changes;
- It was very difficult to quantify the impact of the model, but the level of production quality has increased since implementation. Another way of evaluating the model impact was to analyse the consistency of the operations and the confidence of the users.

After successful first implementation in one plant, three more plants from the company had their own version of the model. Further development included three-point prediction update, full chemical analysis and temperature prediction, dynamic probability plots.

6 References

1. Bhadeshia H.K.D.H.: Neural Networks in Material Science, ISIJ International, vol.39, 1999
2. Delany SJ, Cunningham P, Tsymbal A, Coyle L.: A Case-Based Technique for Tracking Concept Drift in Spam Filtering. 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, 2004
3. Adderley M.: Use of artificial intelligence technologies for process enhancement: Quick-tap phosphorus prediction model, July 1999, TataSteel internal communication
4. Janet Kolodner: Case-Based Reasoning. Morgan Kauffman Publishers, Inc., San Mateo, CA., 1993.
5. McSherry D.: Precision and Recall in interactive case based reasoning. 4th Inrenational Conference on CBR , ICCBR-01, Vancouver, BC, Canada, 2001, Spriger Verlag
6. Lenz M, Auriol E, Manago M.: Diagnosis and Decision Support. In: M. Bartsch-Sporl, H.D.B., and Wess, S.(eds)Case-Based Reasoning Technology:From Foundations to Applications, Springer-Verlag, 1998 LNCS 104
7. A. Aamodt, E. Plaza (1994): Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, Vol. 7: 1, pp. 39-59.
8. M. Adderley, R. Clarke, S. Thornton:Nearest Neighbour Retrieval Software Tool for use in Case Based Reasoning Approach to Problem Solving, August 1999, TataSteel internal communication
9. Saeed, S., Petridis, M., Knight, B. and Mileman, T.: Automating Case Creation and Selection for a Case-Base of Rotationally Symmetric shapes for the Design of Metal Castings, in Proceedings 7th UK Workshop on Case Based Reasoning, Ed. B. Lees, BCS SGES, Cambridge 2002.
10. David B. Leake : Case-Based Reasoning: Experiences, lessons and future directions, AAAI Press/MIT Press, 1996
11. David B. Leake, Andrew Kinley, David Wilson: Case-based similarity assessment: Estimating adaptability from experience, Proceedings of the 14th National Conference on AI, AAAI Press, Menlo Park, CA, 1997