CATMInE
Computer Assisted Trade-Mark Infringement Evaluation

Pierre Renaux
GREYC - UMR 6072
Université de Caen
Caen, France

Abstract. In forensic Machine Learning, information systems are not easy to use due to many constraints. Knowledge generally comes from laws and their use is represented by decisions. The whole data requires several parses in order to obtain a structured database on which decision tools can be applied. This study presents the use of machine learning tools to assist the lawyer in nominative trade-mark infringement evaluation. The domain formalism and database are obtained from the JURINPI database which regroups thousands of cases in the trade-mark domain. The use of a neural network and of Sammon's non-linear mapping provides a development of datas with designed decision tools for trade-mark infringement evaluation : CATMInE (Computer Assisted Trade-Mark Infringement Evaluation). This web-oriented application requires little technical knowledge on machine learning and displays a readable and concise information.

1 Introduction

context : This paper suggests a system dedicated to problem resolution on some machine learning concepts. Trade-Marks represent the company in consumer's mind and thus have important economical issues. There are many registrations and checking for their validity becomes difficult. This work is done by legal advisors for trade-marks owners in order to make the most of their trade-marks portfolio.

motivation : This working context, less studied, represents an important task for lawyers and data is still increasing. Working with it is not complete as it could be for other application areas as business or medicine [2].

The central claim for this research activity is to provide machine learning tools for legal advisors, where behaviour isn't built on imitate human decision making, as expert systems, but rely on case based learning. As far as, applying these tools mustn't imply knowledge on the used machine learning algorithms .

contribution : CATMInE answers to the presented motivations by offering a decision process for nominative trade-mark infringement. This system uses a database and tools where maintenance can be done without the help of the machine learning expert. The whole system is also designed as knowledge repository and provides users an Internet access to the system.

organisation : First, this study presents the nominative trade-mark infringement database and the features extracted to exploit it. Next, pre-hypothesis and their results for knowledge representation in CATMInE are presented.

Pierre Renaux, 'CATMInE - Computer Assisted Trade-Mark Infringement Evaluation' in D. Bourcier (ed.), *Legal Knowledge and Information Systems. Jurix 2003: The Sixteenth Annual Conference*. Amsterdam: IOS Press, 2003, pp. 61-70.

Table 1: Types, tags, and available information used for trade-mark infringement storing.

Information	tags	expressions
Reference	<REF>	*M19950787* or a date *M* = Marque *1995* = year *0787* = order in the year
Jurisdiction	<DOC>	level *place (Court)*, date
Decision's date	<DAT>	*Year-Month-Day*
Parts	<PART>	*Plaintif1*; ...; *Plaintif n*/*Dfence1*;...
Trade-Marks	<MARQ>	*Trade-Mark1*; *Trade-Mark2*
Analyse	<ABS>	*Infringement (YES)*
registration numbers	<OBJ>	*1546621;910560*
Products and services	<CL>	*CL01*, *CL02* *CL01 CL25*
Referenced judgement if it is needed	<LIEN>	*level place (Court), date* or a reference : *M19950787*
Complete decision	<TXT>	*About decision ...*

The third section gives an overview of CATMInE and its functionalities developed to assists the lawyer during a trade-mark infringement search or study. Finally we present the experimentations and a brief discussion on the system's evolution.

2 Application Area

The domain for this research activity is the french trade-mark infringement case-law. It is taken from the online database JURINPI which offers two major interests. First, it is a real database (as opposed to an artificially created one) and second it is partially represented within an XML structure.

2.1 The Trade-mark Infringement Evaluation Task

The trade-mark domain is growing up. For a company, its trade-marks reflect its identity to consumers, for targeted products and services. Having an opponent that uses a similar trade-mark for the same products or services becomes a commercial constraint. The deformed trade-mark could confuse a standard consumer, and the society would loose market activities.

To protect their rights, societies ask for intellectual and industrial property law protection in order to increase the value of their trade-marks portfolio. In case of a trade-mark infringement, the resolution can be done by a court. The taken decision is then inserted into JURINPI.

The JURINPI database represents a part of this case-law[1] and offers an access for consulting, from the Internet, to a legal advisor, for a consideration.

2.2 JURINPI : Database u

The JURINPI database is built on a XML structure to represent decisions. A decision can be seen as a form. The XML structure is composed of 19 tags. Only 10 tags are used for nominative trade-mark infringement. The remaining tags are designed to comply with patent case-law and database indexation. Table 1 describes tags and available information used to represent the nominative trade-mark case-law.

[1]it represents french patent cases law too

In this structure, many elements are difficult to extract. First, in the case analysis, the verdict is not the only information expressed. The information consists of case summaries. Moreover, in nominative trade-mark infringement, two types of verdict are possible : reproduction (total or partial) and counterfeit. The first deals with infringement seizure. The second case corresponds to our targeted research domain. Finding the verdict implies complex mechanisms in order to be extracted.

On the same principle, the "Owner/Trade-Mark", "Trade-Mark/Product and Services", or "Trade-Mark/Registration Number" associations are not easy to establish : enumeration order in the XML structure is not respected for all the tags. The only way to do it is to analyse each case completely in order to validate or to correct the computed associations.

3 Domain Hypothesis

Completing this research activity on the JURINPI database implies the extraction of relevant information to establish a computational structure. The aim is to define linguistic features in order to apply machine learning tools on it.

To establish if a trade-mark imitates another, it is necessary to compare them on their whole aspect in order to find if a risk of infringement exists between them. As the domain is the nominative trade-mark, it is possible to consider a trade-mark as words. So comparing them is exactly as finding common properties on words.

To define a case law, we calculate similarities on the way trade-marks are written, the graphemic aspect, and the way trade-marks are pronounced, the phonemic aspect. These similarities came from an expert know-how and, are calculated, in way of characters, between two trade-marks in the graphemic and in the phonemic representation. So we can establish comparison points on character similarity, phoneme similarity, graphemic string, phonemic string and co-positioning strings for graphemic and phonemic aspect. All these elements came from a large study of case-law, and, from elements used by court to apply laws. These values are next stored in a linguistic similarity vector (one per pair of trade-marks) for future treatments.

For example, if we study the pair of trade-marks (GOLF PLUS, GOLF'US)[2] for graphemic level, we can observe :

- GOLF PLUS is composed of all the characters from GOLF'US, GOLF'US contains 6 characters from the 9 of GOLF PLUS
- GOLF is the longest string common to the two trade-marks
- GOLF is the opening word of the both trade-marks

At the phonetic level, parsing is identical, but applied on the phonemic transcription of the trade-marks pair. This transcription is obtained by using a contextual rules translator. Using a phonemic lexica to translate trade-marks is not possible. Trade-marks are commonly neologisms (a neologism is never present in a lexica). So using a rule based translator is more efficient : easy maintenance, neologism representation and acronyms pronunciation integration are possible.

Another aspect of this research area deals with the different court levels. Applied to french courts, the first court level, Tribunal de Grande Instance is less important than the Cours

[2]Golf House (ste, Italie) versus P. Schmidlin, imitation infringement, Tribunal de Grande Instance, Paris (Ch.03), 01/09/1999

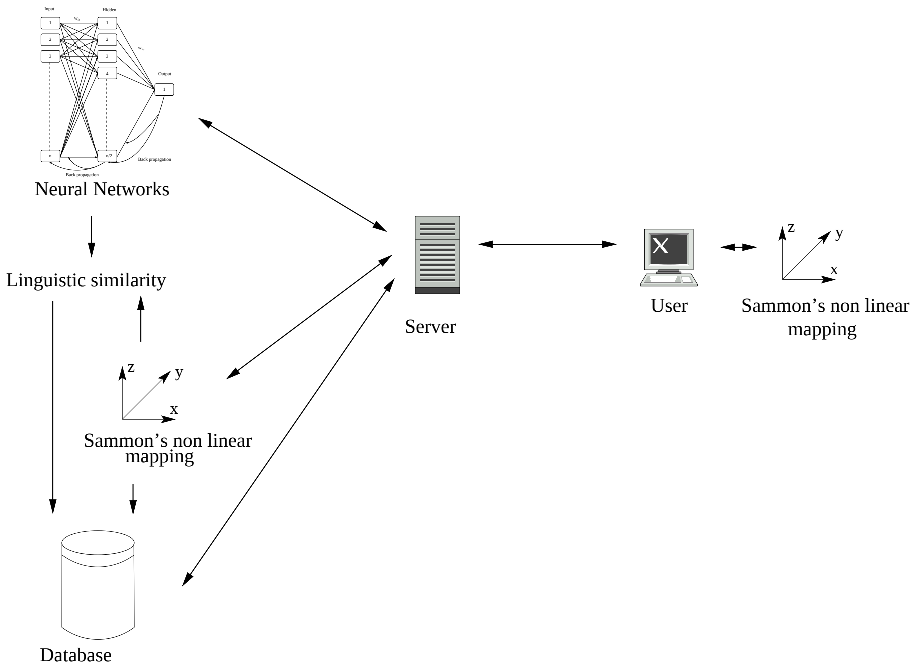

Figure 1: CATMInE's dependencies diagram

d'Appel, which is also less important than the Cours de Cassation. By this relation between levels, the more important the hierarchy, the more the decisions weigh over lower court

This hierarchy justifies the modelling necessity of nominative trade-marks infringements, using case-law. The infringement risk evaluation, needed to prevent law suite between trade-marks, is obtained on a case based learning process, put into practice by a neural network [7]. With the previous definition, having an expert system for this evaluation does not comply with the court's level principle.

4 Computer Assisted Trade-Mark Infringement Evaluation

4.1 Overview

CATMInE software was developed by myself during a Research and Development project between our industrial partner Breese & Majerowicz and the University of Caen by the way of Thomas Lebarbé. It uses the two hypothesis presented before, and can work on any case-based legal system, as the United Kingdom one, just by changing the contextual rules translator and case-base (it also implies a learning process on this case-base). It can be used by the way of a web server application. The designed software is presented in diagram 1, p.64

4.2 First use of CATMInE : Evaluating a Risk of Infringement

Evaluating a risk of infringement between two nominative trade-marks is obtained by computing the linguistic features from the pair of trade marks given in input (see 3). This evaluation depends on graphemic and phonemic similarities. It also depends on the studied country. The phonemic translation and learning database are adapted to the targeted country. Results are then stored in a linguistic similarity vector, one per country.

When features have been computed, it is possible to compare the input case to the whole case-base with the neural network, which has been previously tuned. Result is expressed by

Figure 2: Screen shot of infringement risks evaluation between Velux and Faelux for the french jurisdiction

a score varying from 0 to 100 (near 100, high risk; near 0, low risk). Figure 2, presents this kind of infringement evaluation for the pair of trade-marks (Velux,Faelux)[3].

Moreover, with the linguistic similarity vector, the system is able to display the input case among the whole. This representation is obtained with three proximity case-law mappings. These graphics are drawn using the Sammon's non linear mapping algorithm ([1], [3]) which allows to project a set of points (one per decision) from an n dimensional space to an m one, where $m < n$ and spatial distances are preserved as much as possible. With this algorithm, the software can produce a map of the case-base where the user can observe the surrounding cases in order to study on which grounds the decision has been taken for nearby cases. Observation can also be done on a selected court level or city. Figure 3, page 66 shows the phono-graphemic[4] map (there is one for each graphemic and phonemic aspects) on the french case base. Red cross represent cases where infringement evaluation concludes to yes, blue ones correspond to a negative decision.

By gathering the surrounding cases on the mapping, CATMInE produces a summary of the comparisons by finding the five most similar decision for each mapping. From this list, the lawyer can obtain the text of decision in order to study on which grounds the decisions were taken in order to target his study as best as possible.

[3]V KANN RASMUSSEN INDUSTRI A/S (Ste, Danemark) and VELUX FRANCE (SA) versus S.N.C. FAELUX DI FANTINI SERGIO EC (Ste, Italie), Cours d'Appel, PARIS (CH.04), 25/02/2002

[4]combined graphemic and phonemic features

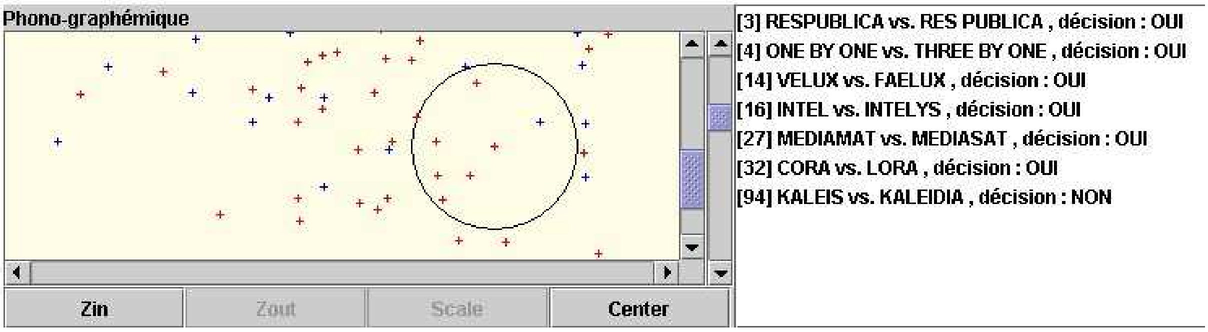

Figure 3: Phono-graphemic map for the (Velux/Faelux) infringement evaluation on the french jurisdiction

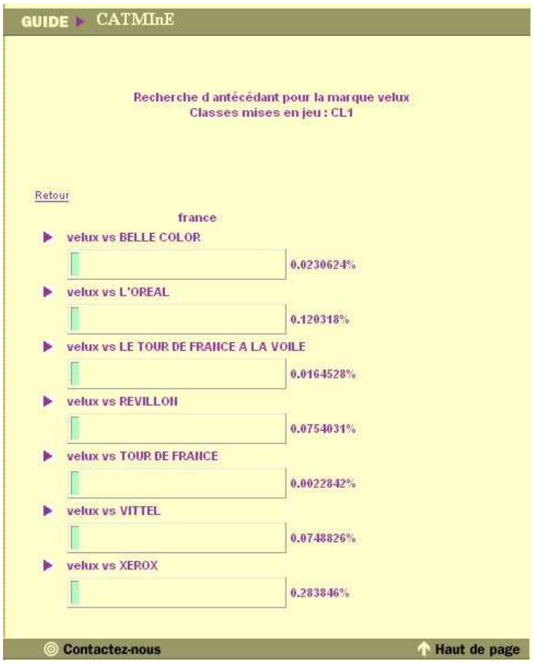

Figure 4: List of antecedents extracted by similarity from the french case base for the trade-mark Velux on the First class of products and services.

4.3 Second use of CATMInE : Sorting a List of Antecedents

Finding a list of antecedents aims to find out if a new trade-mark generates a strong infringement risk with existing ones, for the same or similar targeted products and services. This search usually represents long investigations but is simplified by the use of the CATMInE's database.

By querying the database, it is possible to obtain the whole list of trade-marks for the targeted products and services.

The linguistic similarity vector is computed for each trade-marks pair (composed of the new trade-mark and one from the query) and their infringement risk evaluated by the neural network (this is the same method as the one used for comparing trade-marks). Results of this action is the list of all the trade-marks (sort by name), for the targeted products and services, and their potential infringement risk with the trade-mark given in input.Figure 4, page 66 presents results for a kind of search.

4.4 Third use of CATMInE : Doctrinal use

An interesting point, in this way of using CATMInE, is the evaluation of a recent decision. The neural network tells the degree of similarity with case-law. This provides new tools for the lawyer to study and explain the decision with the mapping.

The second point is to study, globally, the case base evolution through time. If the system is regularly updated with new decisions, it becomes possible to have a look on decisions given by a court. It is also feasible to study jurisdiction's sudden changes : finding forewarning elements in order to anticipate them by observing the Sammon non linear mapping.

5 Discussion

CATMInE corresponds to new tools in the domain of Artificial Intelligence applied to Law. From this research in the domain of computer assisted decision process, some problems emerged and are presented here.

The first problem is of a semantic order. The aim is to use in the decision making the semantic similarity between trade-marks. For example, if we study the trade-marks pair "new man" and "new boy", the surface linguistic similarity wouldn't argue to infringement. But in this case there is infringement due to semantic similarity between "man" and "boy". A way to solve this problem could be a user specification during the input of the case.

A second problem is the automatic language detection. For registering a new trade-mark, it is possible that the rules for translating the trade-mark into its phonemic equivalent are not adapted. To illustrate, it is possible to register an english trade-mark in France (and in the same way a french trade-mark in England). The system will use the french's transcription rules to produce the phonemic representation. So, this transcription might be false : the final 's', for example, wouldn't be transcripted and may change the phonemic similarity with the other trade-marks. This constraint implies that the legal advisor defines which rules the system should apply for each word of the trade-mark.

The third problem is based on the length of the compared trade-marks. If one is bigger than the other, the system should integrate features in order to reveal it during the learning process.

Last, the fourth problem refers to the neural network which could be replaced by another decision process in order to explain the elements which conclude to an infringement and trace the decision process. This point refers to my research activities and is briefly presented in the next section.

6 Evolution

The previous section highlights that a neural network could not explain its decision. The hypothesis was to replace it by another decision process as decision tree.

C4.5rules[6] was the first tested algorithm, with defaults parameters. Its results were not good enough. It generally produces 1 rule and a default one, for sometimes the same targeted decision. Its average error rate is around 35%. Experiments show that c4.5 usually use the default rules to classify a new example. In this way, decision explanation is not powerful.

The next section presents a part of my research activity : quality measures applied to decision tree in an further version of CATMInE, named DETMInE.

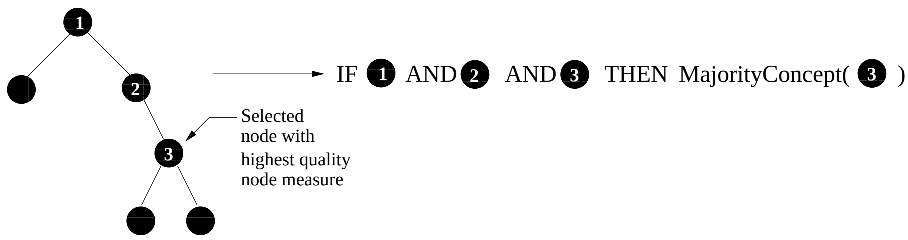

Figure 5: rule construction.

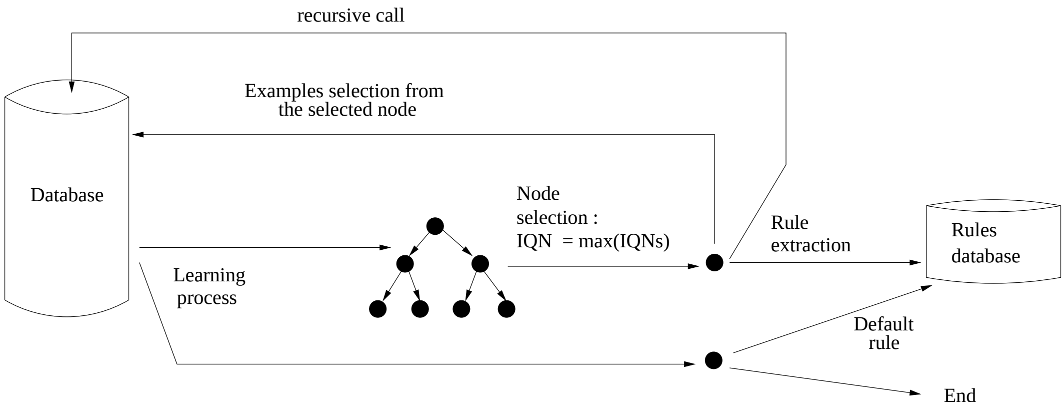

Figure 6: Recursive algorithm for rules selection.

6.1 Decision Tree for Trade Mark Infringement Evaluation

The working hypothesis was to classify majority examples with simple rules and complex examples with complex rules. So, by using a quality node measure on decision tree[5], the hypothesis can be developed. This quality node measure depends on three parameters :

1. the weigh of the node in the tree (number of associated examples to the node)
2. an impurity measure normalised between 0 and 1
3. a damping function which depends on the depth of the node in the tree

First, the decision process builds a decision tree form the learning database. Next it searches the node which presents the highest quality node measure. A rule is then produced from the selected node as presented in the diagram 5, p.68. The associated examples are deleted from the learning database and the highest court's level is extracted from the deleted examples. The court's level is then associated to the rule as an extended information.

Last, the algorithm rebuilds a new decision tree from the rest of the learning database and extract other rules in the same way, until the decision tree is reduced to a root node. This root node will then represent a default case to apply if no rule is available.

This algorithm is summarised in the diagram 6, p.68.

When a new example is studied, the court level will be used in case of conflicts between rules : if two rules are available for this new example, the algorithm selects the rule which presents the highest court's level. If available rules have the same level, the selection is then made on their highest level's population.

Table 2: Ten-fold Cross Validation's error rates.

Experiments	Neural Network Accuracy	DeTTMInE Accuracy
Experience 1	60%	60%
Experience 2	80%	70%
Experience 3	90%	80%
Experience 4	70%	70%
Experience 5	70%	70%
Experience 6	70%	80%
Experience 7	80%	60%
Experience 8	70%	70%
Experience 9	70%	60%
Experience 10	60 %	70%
Average	72%	69%

6.2 Experimentation

The goal of these experiments is to compare the neural network accuracy to the DeTTMInE's one. Their accuracy is estimated using a ten-fold cross validation([4]). Accuracy of the learned classifiers are summarised in the table 2. In the overall, the neural network seems to be slightly better than the decision tree algorithm for only 3%. But the induced rules, by the new algorithm, offer explanations to lawyer in many ways. First it can produce all the examples that satisfy this rule (examples extracted during the learning process). Those examples can be used by the lawyer for a better understanding of the studied example. It can also show a rule, and so features that are relevant for the suggested decision.

Another important result is related to the number of produced rules and their use. In opposition to c4.5, DeTTMInE produces around twenty rules and uses around fifteen of them. So, explanation of new examples is not justified with a default rule, but by a set of rules that satisfy the objectives of explanation and information needed by CATMInE for the decision making.

7 Conclusion

The presented system, CATMInE, uses machine learning algorithms to provide support for trade-mark infringement evaluation. First a neural network for a quick infringement evaluation between two trade-marks, and next a N-dimensional case based mapping for observing the surrounding cases and trace the decision process. The evolution on the decision process offers the ability to explain a comparison or antecedent search without changing the application and represents a starting point for further research.

The guidelines for this work allows a legal advisor to study infringement risk between trade-marks without machine learning expert. The effectiveness of the comparison/visualisation system was tested and validated onto two case-bases : a french one and an english one.

CATMInE was designed for our industrial partner Breese & Majerowicz. Available on the Internet, it allows a knowledge repository for available country. The multi-lingual aspect supports infringement evaluation on any available country, European community and maybe more.

Professionals who use CATMInE, show the interest to have simple and explanatory tools at their's disposal, able to assist them daily and was demonstrated in May 2003 at the International Trade-Mark Association annual meeting (INTA'03).

References

[1] J.W. Sammon, A nonlinear mapping for data structure analysis, IEEE Transactions on Computers, vol. 18, pp 401-409, 1969

[2] S. Ivkovic, J. Yearwood and A. Stranieri, Visualizing Association Rules for feedback within the legal system, ICAIL03 proceeding, pp 214-223, 2003

[3] Biswas G., Jain A.K., and Dubes R.C., Evaluation of projection algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 3, pp 701-708, 1981

[4] Petrak J., Fast Subsampling Performance Estimates for Classification Algorithm Selection, ECML, Workshop, 2002

[5] Fournier D., Thèse, Etude de la qualité de données à partir de l'apprentissage automatique. Application aux arbres d'induction, Université de Caen - Campus II, 2001

[6] Quinlan J. R., C4.5: programs for machine learning, isbn : 1-55860-238-0, publisher : Morgan Kaufmann Publishers Inc., 1993

[7] McCulloch W.S. and Pitts W.H., A logical calculus of the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, vol. 5, pp115-133, 1943