The role of AI in evidence-based strategic IP decisions

As patents increasingly take shape as a functional asset class, the use of artificial intelligence looks set to lead to more licensing and less litigation

By Nigel Swycher and Steve Harris

Patent strategy used to be synonymous with the preparation and prosecution of patents. Success was measured by the number of patents filed, while benchmarking involved a company comparing aggregate numbers against those of its competitors. However, such days are long gone, replaced by reporting that requires a company's IP strategy to be aligned with its overall corporate objectives. This evolution has increased the demand for greater transparency and evidence-based decisions.

An evidence-based approach focuses on a consistent and repeatable methodology that can clearly demonstrate how patents map to a specific technology and how they relate to all other patents for the same technology. This calculation is already used, explicitly or implicitly, by teams responsible for patent licensing. It is also finding favour in litigation, where courts are tasked with resolving disputes involving large global portfolios.

However, there is no practical way for humans to count and sort large populations of patents. Not only are there too many patents but it is rare for two experts to agree on the other's approach to mapping patents to technologies. This article explains how artificial intelligence (AI) – specifically binary classification – can be used to increase accuracy. In addition, it looks at how established techniques for measuring precision and recall can reduce the amount of disagreement in relation to results.

Having reduced the general case to an equation and explained how binary classification can be used to populate key variables, the approach is then applied to a range of strategic IP decisions where there is an increasing demand for evidence-based decision making.

Making money from patents

The analysis begins with the common case. One party (*a*) owns a large patent portfolio that relates to multiple technologies (product lines). It is aware that companies are likely to infringe many of the patents and develops a monetisation strategy aimed at maximising economic return. At the heart of the evidence required to hone the strategy is the following formula:

$$\text{Mean royalty rate: } s_{pa} = s_p \frac{n_{pa}}{d_p} \qquad \text{Equation 1}$$

For a given product line p and company a where:

- n_{pa} is the number of patents owned by a that read on to p;
- d_p is the total number of patents reading on to p, regardless of owner; and
- s_p is a royalty rate typically attributed to p.

Ideally this calculation would be carried out for each country individually, with separate patent counts. However, in practice a useful estimate can be reached by calculating at the patent family level and weighting for revenue distributions across different countries. Where companies have a significant skew between the geographical coverage of their patent portfolio and revenue sources, it is necessary to carry out the calculation for each region.

While it may be controversial to think about patent licensing from an economic perspective, Equation 1 is self-evidently true. If s_p is a representative royalty rate, then a party that controls all of the patents reading onto the product and has licences with all parties producing that product would, by definition, receive s_p × the revenue from that product.

Of course, this situation never occurs in reality. However, apportioning the hypothetical global licence over multiple parties gives the formula above. It only indicates the mean royalty – not the distribution. So while it cannot be used to directly price licences, it can be used to derive a number of useful metrics.

The obvious challenge to this estimate is that not all patents have equal value. This is clearly true but there is plenty of evidence – both empirical and theoretical – that for any particular application and within any particular portfolio the value of patents falls along a Pareto distribution (eg, Putnam J, "Value Shares of Technologically Complex Products" (2014), SSRN, and Harrison, S, Sullivan, P, *Edison in the Boardroom Revisited: How Leading Companies Realize Value from Their Intellectual Property*, 2nd Edition, pp 31-33, ISBN: 978-1-118-00453-1).

The logical conclusion of this is that in any mature product area there will be a relatively small number of high-value patents and it is highly unlikely that these will be concentrated in a single portfolio. The remainder will largely be of similar value to each other and the distribution of value will be similar across all substantial portfolios.

This also raises the question of: valuable for what? Depending on the specific issues faced by a specific company some sets of patents may be especially applicable to their needs but will not be equally valued by other players.

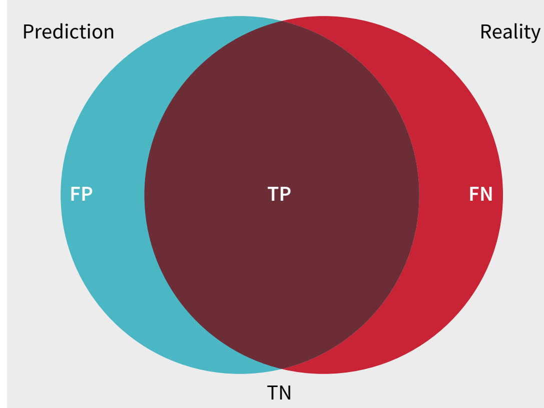
So while it is possible that a company's own portfolio (or that of someone seeking licence revenue from it) might contain a large proportion of all the valuable patents in a given area, it is highly unlikely.

The difficulty of applying these metrics in practice comes down to determining the variables in the equations. This article focuses specifically on techniques for determining *n* (the numerator of the proportional fraction in Equation 1) and *d* (the denominator). An illustration of how this works in practice can be found in Judge Selna's decision in *TCL v Ericsson* (see box-out). While there is disagreement about whether national courts should attempt to determine global royalties, it is easy to appreciate why there is a trend in this direction.

The starting point is simply practicality. Markets are global and patents are national. No one (except perhaps patent litigators) believes that it is efficient for there to be multiple litigations all relating to the same fundamental dispute. Disputes are increasingly based around portfolios (or at the very least technologies) and litigation typically requires the plaintiff to focus on a few claims and a handful of patents. There are also many situations where the resolution of the dispute will require an assessment of claims going the other way (as illustrated by the extended formula: Equation 3).

In his judgment in *TCL v Ericsson*, Selna noted that "the search for precision and absolute certainty is a doomed undertaking" – reflecting the reality that litigation is not an ideal forum for running these calculations. This is where algorithmic analysis can help. However, advances in AI and machine learning can improve the calculation of *d* and *n*. Such an approach would be a significant improvement over current manual approaches.

FIGURE 1. Precision/recall



Binary classification primer

The task of establishing whether some object belongs to one of two mutually exclusive classes is known as binary classification and is one of the most studied problems in computer science. In this situation the object is a patent family and the classes are patents that read onto a particular product line and those that do not.

Applying a binary classifier to a patent family will produce a result that allows the system to put the family into one of two classes: positive (ie, the family reads onto the product) or negative (ie, it does not).

There are also many non-AI algorithms capable of performing binary classification. Some of the more common ones are logistic regression, Naive Bayes classification and support vector machines.

Quantifying accuracy

In order to make meaningful statements about the relative accuracy of different classification techniques (algorithmic and manual) it is necessary to quantify accuracy.

Conventionally this is done using two metrics, precision and recall (see Figure 1). Precision is the proportion of answers returned which are correct, and recall is the proportion of correct answers in the input set which are identified. In order to calculate precision and recall, the classifier is applied to a set of patents where the classification is known, but where none of those patents have been used to train the classifier. Each result can then be divided into one of four groups:

Group	Prediction	Reality
True positives	Positive	Positive
False positives	Positive	Negative
True negatives	Negative	Negative
False negatives	Negative	Postive

From that, precision is defined as $\frac{tp}{tp+fp}$ and recall as $\frac{tp}{tp+fn}$.

TCL v Ericsson: a case study

Ericsson owns many SEPs relating to 2G, 3G and 4G cellular technologies. TCL manufactures and distributes cell phones globally. The parties failed to agree licensing terms. As part of the process of determining the royalty due, the court used this approach:

Total aggregate royalty × Ericsson's proportional share = Ericsson's royalty rate

Where :

Ericsson's proportional share = $\dfrac{\textit{Number of unexpired 4G patents owned by Ericsson}}{\textit{Total number of 4G patents globally}}$

This is equivalent to Equation 1, where s_p is the "total aggregate royalty", n_{ap} is the "number of unexpired patents owned by Ericsson" and d_p is the "total number of 4G patents globally".

The focus is on the calculation of the numerator (SEP patents owned by Ericsson) and the denominator (total number of SEP patents globally) and ignores distinctions between 2G, 3G and 4G patents.

Predictably, there was disagreement in the value of both *n* and *d*. TCL argued for a low value for *n* and a high value for *d* (which would result in a low percentage of the aggregate royalty), and Ericsson was motivated to argue the exact opposite.

For *d*, the starting point was the 153,000 patents (including applications) declared to be essential. There are two big problems. Not all declared patents are essential and not all essential patents are declared. TCL employed consultants, including Ernst & Young, India, to conduct a study to determine *d*. Inevitably Ericsson challenged the results from every angle.

First, the quality of analysis. In order to grind through the pile, it was calculated that the TCL team spent 20 minutes per patent family. It is entirely reasonable to question whether a human can map a patent onto a detailed technology specification in that time. A related criticism was whether the people who conducted the review had the level of technical expertise required. Second, bias. The teams retained by TCL knew which side they were on and the answer that would best suit that client. For *n*, Ericsson put forward over 190 patent families and TCL disputed over 60 of them. Other areas of disagreement included:

- geography – TCL sell products in the United States, Europe and Asia. Ericsson's patent coverage was less outside the United States and the court accepted that its royalties should be calculated regionally, using the reduced coverage as a scalar; and
- strength – TCL argued that Ericsson's numerator should be adjusted for patent strength on the basis that weak patents should count for less. This assessment was part manual and part driven by analysis of forward citations. The court held that TCL's analysis was fatally flawed.

TABLE 1. Battery technologies owned by leading Japanese automotive companies (active patent families)

Technology	Toyota	Honda	Nissan	Mitsubishi	Suzuki	Mazda	Subaru	Isuzu
Battery management systems	1,558	276	297	144	78	51	59	24
Battery separators	226	87	78	0	0	0	1	0
Battery thermal management	644	137	154	121	54	40	43	9
Lithium-ion batteries	1,268	24	310	36	0	1	12	0
Lithium-oxygen batteries	34	18	2	3	14	0	0	0
Lithium-sulfur batteries	38	0	6	1	0	0	2	0
Metal-air batteries	117	10	73	1	6	0	0	0
Nickel-metal hydride batteries	55	2	0	0	0	0	0	0
Solid state batteries	428	0	20	2	0	0	0	0
TOTAL	4,368	554	940	308	152	92	117	33

A hypothetical perfect classifier would have a precision and recall both of 1.0, but in reality – and especially in the patent world – the membership of some patents to some class is highly subjective. In such cases, the best that can be done is to compare the precision and recall against some 'gold standard', results which are believed to be representative in some way.

There are a lot of ways to estimate the precision and recall of a classifier, but one of the most common is k-fold analysis, where a portion of the training set (ie, the patent families used to train the classifier) is withheld and another similar classifier is trained on the remainder. This is repeated for every portion of the training set withheld (eg, if it is one-tenth there will be 10 folds), with the precision and recall of each fold then analysed to estimate the precision and recall of the parent classifier.

This is only ever an estimate, as the training set is typically not representative of the whole world, reflecting only families of interest to the classifier. However, it can result in good estimates if applied carefully.

Why use classification, not sampling?

While machine classification will never be as accurate as an expert spending hours poring over every patent, it has the advantage of repeatability and scalability. A sophisticated AI classifier can determine the class of a patent family in milliseconds, whereas an expert will require minutes to come to the same decision.

The use of experts to estimate sizes of populations of patents is especially problematic when it comes to estimating denominators. Suppose a situation where there are 100,000 families in the plausible domain of the task (eg, ones which include CPC codes known to relate to the product), where only 1.2% (1,200) of these truly relate to the product – perfectly plausible numbers in the real world. It would require 4,000 of these families to be read and classified by the expert, just to get a 95% confidence interval range of 898-1600 for the size of the denominator (the Agresti-Coull Interval, Agresti, Alan; Coull, Brent A (1998), "Approximate is better than 'exact' for interval estimation of binomial proportions", *The American Statistician*, 52: 119-126).

In comparison, a classifier with a mean predicted precision and recall of 0.89 with standard deviation of 0.034 (results which are easily achievable in Cipher) will achieve a 95% probability range in the region of 1,067-1,333, a 2.6x narrower interval. Deriving probability ranges from analysed classifier characteristics is a complex topic and outside the scope of this article.

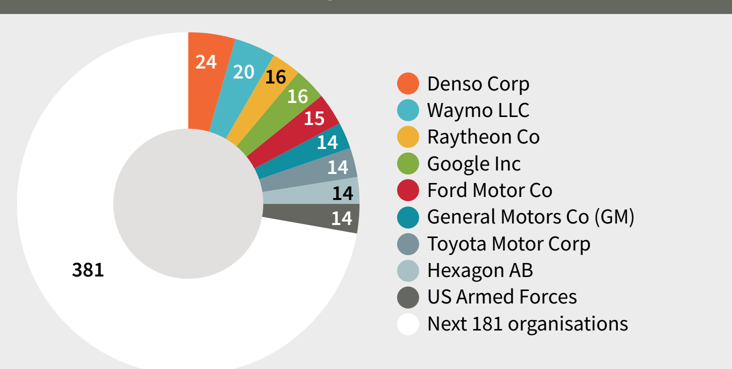
The resources required to get expert opinions on 4,000 families (around one person-month, assuming a mean of 2.5 minutes per family) would be excessive, even for very large problems. In addition, the accuracy of the end results would not be sufficient for many purposes. Indeed, in order to match the accuracy of the hypothetical classifier, a perfect human expert would have to sample 27,000 families, spending around 6.5 person-months of effort.

Because of this, an inferior classification process applied exhaustively over the population will generally achieve a higher level of accuracy than a theoretically perfect one applied to a relatively small sample. An AI classifier, trained with the order of 400 patents (around two person-days of effort), can easily achieve the same precision and recall characteristics as the example above – a substantial saving in resources. In addition to that, once trained, the classifier can be reapplied again whenever it is needed, rather than this being a one-off manual task, which must be repeated for each analysis.

Putting theory into practice

AI classification of patents is already used in commercially available software – an example being an analytics platform that is built around the taxonomy

FIGURE 2. Automotive LIDAR families with US grants



Equation 1 can be generalised to estimate the royalty revenue achievable by company *a* to a product line *p*, for a set of monetisation targets, *T*:

Mean royalty revenue, $r_{paT} = s_p\, r_{pT}\, \frac{n_{pa}}{d_p}$ Equation 2

Where:
- *rpT* is the total revenue earned by all companies in *T*, attributed to *p*;
- *npa* is the number of patents owned by a which read on to *p*;
- *dp* is the total number of patents reading on to *p*, regardless of owner; and
- *sp* is the royalty rate typically attributed to *p*.

It can also be extended to estimate the balancing payment in a cross-licence agreement, from company *b* to *a*: *bpab=rpab-rpba*, so:

Balance of payments, *bpab* = *sp* $\left(\frac{r_{pb}\, n_{pa} - r_{pa}\, n_{pb}}{d_p}\right)$ Equation 3

Negative results indicate that the balancing payment should go from *a* to *b*.

of over 200 automotive technologies. Table 1 is a visualisation of a range of battery technologies owned by a selection of automotive companies. This is the algorithmic calculation of a range of (*n*s) with a precision and recall in the region of 0.9.

Compelling reasons to know who owns what

There is a compelling need for evidence to support a broad range of strategic IP decisions.

Cross-licensing: the balance of trade

The basic formula resolves a one-way transaction: where one party (*a*) owns patents and is asserting that the other party (*b*) is infringing. However, more often than not, *b* has a reciprocal claim that *a* infringes its patents. In such situations, the balancing payment is represented by a logical extension of the base formula (see Equation 3). This formula calculates the counter-claim from *b* (applying its *n* against *a*'s revenue) as an off-set against the royalty likely to be generated from *a*'s patents when applied to *b*'s revenue. To apply this formula requires knowledge of a typical royalty rate – often known to the company – and estimates for product-specific revenue.

Using the data set out in Figure 2, a hypothetical cross-licence estimate could look like this (where patent numbers are accurate but revenue figures are arbitrary):

Company	Families (n_{px})	Product revenue (r_{px})
Waymo (*a*)	20	\$14.5 million
Raytheon (*b*)	16	\$31.3 million

The denominator d_p is 528 (the sum of all holdings). For the sake of example, assume that industry knowledge suggests that 7% (a factor of 0.07) is a typical royalty. So:

$$b_{pab} = s_p \left(\frac{r_{pb}\, n_{pa} - r_{pa}\, n_{pb}}{d_p}\right)$$

$$= 0.07 \left(\frac{31.3 \times 20 - 14.5 \times 16}{528}\right)$$

$$= 0.052$$

Which suggests that in this hypothetical situation Waymo should receive royalties of approximately \$52,000 per year as the balancing payment.

Monetisation: sale versus licensing

'Monetisation' is the generic term applied to the licensing and sale of patents, where the activity is ancillary to the core business. Companies approach monetisation with three primary questions:
- What to monetise?
- Who to target?
- How to maximise return?

Classification can help in all these areas.

Selecting which patents to monetise is a non-trivial exercise in the face of competing views and the demands of the business. Even where there is an internal taxonomy, it is commonly applied early in the application process and not reviewed, as the claims are amended before grant. There are many situations where technologies evolve in such a way that the patent has relevance in new and different areas to those originally known to the inventor or the patent attorneys supporting the process.

Classifiers overcome these issues as they can be applied retrospectively to a portfolio to establish what patents relate to what technologies, and to establish the estimated market share. This combats the instinct to equate having a large number of patents to the view that it represents a large share of the market (which requires knowledge of the denominator).

Classifiers can be used not only to estimate the value for *d* but also to identify owners. This is effectively a longlist for who might be interested in acquiring rights in the technology area under review. Once again this is not to suggest that companies with small holdings are good targets or the opposite. This will require the input of additional information, particularly revenue.

Deciding whether to sell or license is usually a question of which option will maximise returns. This requires a comparison between the one-time sale price, with the ongoing return from a licensing campaign (with various adjustments such as maintenance costs and discounts for risk of invalidity or obsolescence).

Classifiers can play an important role in modelling these scenarios. The starting position in each case is the estimate of *n* and *d*, that enables the owner to establish its share of the patented technologies. It will also be possible to model the global revenues attributable to that technology and the relevant royalty rates. With these variables, it is possible to analyse the various scenarios in which licensing will generate higher revenue than outright sale.

Returning to the LIDAR example, using Equation 2 results in a prediction that the upper bound of the mean licencing revenue per family is in the order of 0.07 × $\frac{1}{528}$ × the revenue for automotive LIDAR. If revenue is predicted to be about \$3 billion over the next five years (assuming that accounting practice in the company is to look at returns over five years and the patent has five or more years of life left), that gives a mean total addressable market (TAM) licensing revenue for a single patent of around \$398,000 over five years. Assuming that there are no other reasons to retain that patent (eg, if the company is pulling out of that product area), any sale price that even gets close to that range would make it economically sensible to sell.

In addition, it is feasible to do a serviceable obtainable market (SOM) analysis by looking at companies that the company conventionally licenses technologies to and have a reasonable expectation that they would wish to obtain a licence for that technology and use their revenue to calculate the SOM. So, if the revenue from LIDAR sensors for companies in the serviceable available market (SAM) is $100 million, that suggests that the company should not normally entertain selling the patent for less than $13,000 ($0.07 \times \frac{1}{528} \times$ $100 million).

Clearly this range ($13,000 to $398,000) is too large to be applied as it stands. However, by adding in companies that are outside the SOM but from the TAM, it is possible to derive a representative estimate of the likely market SAM. Even simpler methods (eg, the weighted mean of the SAM and TAM) can be used in the absence of other financial data.

If there is a view that with some appropriate effort the company could obtain licences from 80% of the SOM and 10% of the remaining TAM, an estimate of the mean plausible licensing revenue can be calculated as $48,600 per patent ($0.8 \times 13+0.1 \times (395-13)= 48.6$).

In many respects, monetisation is similar to cross-licensing. There are many situations where a critical part of the analysis is an assessment of the potential counterclaim. There are many war stories where patent owners have suffered by asserting their rights in one area only to find that they have a much greater exposure in another area. While this risk is known, the ability to classify and compare portfolios objectively, repeatedly, quickly and at low cost should have a dramatic impact on how these decisions are made going forward.

Acquisitions: patent value

Somewhat different to large companies with monetisation strategies is the patent brokerage market, which supports the many companies that are looking to sell patents and also those who believe for whatever reason that they are under-stocked (eg, Uber's UP3 initiative).

A typical characteristic of the market is that the assets (or lots) tend to be small with owners who believe that they are best in class (ie, high quality). Classification provides an additional perspective to potential purchasers. It enables the lot to be seen in the context of n and d, highlighting the typically small percentage of the total market represented by the lot and the royalty that can be commanded for that technology. While there will always be those who assert that they are entitled to 50% of the royalty for 5% of the rights, classification provides, at the very least, an objective counter balance.

Patenting: what to protect

"How many is enough?" is a question that management always asks of patent teams. The answer is both contextual and relative. The teams responsible for patenting work to a budget, usually derived from the money made available in the previous year. The budget must be allocated to maintenance (of granted patents), prosecution (of pending applications) and new (fresh inventions). It is not uncommon for maintenance and prosecution to consume over 70% of the budget.

A common approach is to try to balance the revenue of the company with the spend on the patent portfolio. However, this ignores the critical factor that the density of patents varies substantially from technology to technology, such that in order to protect the company's position in the market, it is generally better to try to match the company's patent position to its market share.

For example, if a company has a 10% market share in a particular technology, it would ideally hold 10% of the granted patents in that area, such that:

$$\frac{n_p}{d_p} \approx \frac{\text{company's revenue from } p}{\text{all revenue from } p}$$

Ideally this would be done at a country level, as the market share in the United States could be significantly different from China, for example, and the approach supports more detailed modelling of that sort also.

Trusting AI with patents

There are a number of limitations when it comes to using AI to classify patents. First, there is no expectation that the parties will plug and play – that is to say, use the algorithmic computation to impose the outcome of negotiations. What the ability to run these calculations does achieve is to eliminate indefensible assertions of the sort that lead parties to conclude that a negotiated outcome is impossible.

Second, tools that help people to achieve rational outcomes assume that this is the desired objective. There are many situations where litigation, understandably, is the weapon of choice; examples include NPE litigation, where the business model focuses on disproportionate enrichment in the absence of any relevant counter-balancing force.

What is attractive about this approach is its objectivity and rationality. The inspiration comes in part from the Black-Scholes model developed in the 1960s for pricing European put-and-call options. The model was widely adopted, notwithstanding the fact that it makes a number of assumptions about the assets and the market that were always open to a range of views.

While this analysis focuses on the simple case of one party licensing patents to another, this article demonstrates that such an approach can be extended to a range of other strategic IP decisions. iam

Nigel Swycher is CEO and **Steve Harris** is chief technology officer of Aistemos, London

Action plan

A

Consider how best to integrate evidence-based decision making into your IP strategy.

There is a measure of agreement as to the formulae that can be applied and established computer science techniques that can assist with population of key variables. The base formula for licensing as used in cases such as *TCL v Ericsson* can be extended to a range of other strategic IP decisions including:

- cross-licensing, to assist with balance of trade calculations;
- monetisation, to compare outcomes of patent sale versus licensing;
- acquisitions, to provide objective evidence of value for patent purchases; and
- patenting, to help align portfolios to revenue and market share data.

For those sceptical of evidence of this sort, take comfort in the fact that counting and sorting data is what AI algorithms do best. Applying these techniques to help make strategic decisions can dramatically reduce time spent performing laborious and repetitive tasks, and provide valuable insight into your position relative to others.