Dispelling False Positives Myths

Vincent White

In the world of KYC watchlist screening, everyone talks about false positives. Practitioners bemoan their prevalence, and screening software providers claim their solution uniquely tackles and reduces them.
It is meaningless to claim a particular algorithm or method of matching reduces false positives by 'x' percent. The output of course depends on many factors.

When assessing the effectiveness of screening software (which are based on matching algorithms and other techniques) we measure their *precision* and *recall*; their success at identifying true positives, or the inverse, not missing any (false negatives), and not producing significant volumes of false positives.

The first part is straightforward – find all the true matches, don't miss any. But why do so many false positives occur and why they so bad?
They are not necessarily a sign of a less effective screening system – but they have come to reflect a malaise with conventional alphanumeric watchlist screen as it exists today. While there are other determining factors relating to the quality and characteristics of the matched data, the volume and proportion of false positives in a given system is better understood as a measure of efficiency.

What are the causes?

Volume of names being screened against
- Matching against larger datasets will produce more results – of every type.

Intrinsic characteristics of names
- Homogeneity (common names)
- Uncertain name syntax
- Translation of names (legitimate: through naturalisation; illegitimate: changing identity)
- Transliteration from other language script

Secondary identifiers
- Absent, partial or imprecise, or incorrect

In an effort to address these linguistic and name structure uncertainties (in an effort to avoid missing true matches) some form of fuzzy matching is applied, or hard-wired rules that define equivalence between names – this is typically to address linguistic anomalies, such as irregular short-form and nicknames, and nuances that cannot be identified by:

- Phonetic comparison (similar sounding)
- Orthographic comparison (similar spelling)
- Morphological comparison (similar structure)

The net result is to increase the volume of potential matches, and therefore also the number of false positives. It is important to understand and recognise that a large proportion of the high volumes experienced in a KYC screening process are caused by deliberately applying fuzzy matching – this happens because of risk aversion:
We would rather live with larger volumes of false positives than miss a true positive match.

The single most important factor that leads to high volumes of potential results is that there are many individuals who share common names. We rely on secondary identifiers in an attempt to *disambiguate* – that is, to distinguish people with the same or similar name who likely have different secondary identifiers, such as their date of birth.
While this is a practical approach, it is flawed – simply because dates of birth are not always known – even the best alphanumeric risk databases lack dates of birth for at least half of all individuals, and more often it is far greater.
Let us also not forget false negatives. When so much effort is made to avoid them with all the fuzzy matching techniques, why are some true matches missed?
The main reason is not due to shortcomings in algorithms, but rather because of falsification of identity. Once an individual has successfully acquired a new identity, with a different name, date of birth, even nationality, the only way a false negative can be avoided is if that new identity and alias is already known (that is, in the public domain).
And, of course, it is usually not.

Biometric matching

Biometric matching comprises several different types of physiological comparison, including fingerprinting, ocular-based (iris and retina), facial recognition, and DNA.
Each of these types boasts superior precision and recall compared to matching alphanumeric labels because of inherent (near-)uniqueness.
In simple terms, we can say that there is more information in these biometric attributes than in a name or other personal identifier. They are more 'defining' identity characteristics and, unlike names, cannot be changed - at least, significantly, or to a degree that would avoid recognition.

With respect to faces, there are roughly ten people in a global population of 7.5bn, with very similar features ('lookalikes') to our self, that would lead to difficulty in differentiation in facial matching. This is a miniscule proportion compared to identical twins (approximately 0.3% of births globally) who, perhaps surprisingly, can often be successfully distinguished through facial recognition.

As with all forms of matching, biometric matching still experiences false positives and false negatives, however, their occurrence is significantly improved in comparison with name-based matching. Match scoring is based on statistical probability (confidence level) rather than alphanumeric systems which have arbitrary scoring. Both are measuring the 'closeness of match' – the fundamental difference is that with name-based systems closeness of match is not equivalent to the probability of a match being true, and this is largely due to not taking into account relative name frequency.

To give a concrete example:

- Two individuals sharing the same common name might generate a high matching score in an alphanumeric screening system, but not be the same individual.
- Two individuals sharing the same face will generate a high matching score in a biometric screening system, and are almost certainly the same individual.

The most important determinant for the volume of potential matches (and hence false-positives) returned is the threshold setting for displaying matches. In alphanumeric systems this sometimes can be configured at different levels corresponding to risk type (e.g. a lower threshold setting for matches against higher risk data such as sanctions) which helps a little, but nevertheless, homogeneity in names in combination with missing secondary identifiers is still overpowering, leading to many potential results that require further review and investigation.

With biometric screening, the threshold can be set at a very high confidence level, since those cases that are 'somewhat similar' are extremely unlikely to be true positives. Essentially, the threshold level is far more effective filter for discriminating 'maybe' from true or false.

Digital Onboarding

Biometric identification has become commonplace with the advent of smartphone technology. This has accelerated the trend towards digital onboarding, as identity verification can be fully digitalised to deliver a rapid, frictionless initial stage of KYC. The COVID crisis has further compelled firms to cater to increased remote onboarding and transacting. Biometric 1:1 comparisons can be carried out to verify identity with superior speed and accuracy, although remote onboarding also presents new risks and vulnerabilities in the authentication process.Watchlist checking, a 1:N comparison, is a logical extension of biometric identification. In practice, possessing an image of a prospective or existing customer, is a prerequisite. Where an image is not captured as a 'selfie' in a digitial process, firms will still typically obtain a copy of a customer ID document to conduct identity verification. Therefore, the process of biometric watchlist screening can be carried concurrently, at the earliest stage of onboarding.
Combining identity verification with risk screening against watchlists to identify individuals of heightened risk is not only faster and more efficient; importantly, it can also identify cases of identity fraud, where an individual may successfully have passed identity authentication but a conflicting watchlist alert indicates another identity.

Conclusion

Biometric screening is clearly suited to the purpose of screening as part of AML/CFT efforts, but if it is so much better then why have we been using name-based systems for the last 20 years? There are three simple reasons: only in recent years has biometric technology advanced to the point where it is far superior, and it takes some time for all emerging new technologies to get traction and become established; secondly, we now have the means to compile the underlying dataset of facial profiles of high-risk individuals in order to match against; and finally, while facial recognition has reached the stage of mass adoption in identity verification (from unlocking phones to airport security) there are lingering concerns, unfortunately based on misconceptions, about bias and inaccuracy, that have inhibited the evolution to a smarter approach to KYC screening.