



NETSPARK

ポルノグラフィーのアルゴリズム比較

Netspark Ltd.は、10年以上にわたり、テキストや画像、動画に含まれる有害なコンテンツを識別する高度なアルゴリズムを開発し、ペアレンタルコントロール・アプリケーション(Canopy、Netspark mobile)、ISPソリューション(現在イスラエル国外にも展開中)、ソリューションに組み込んでいます。 ペアレンタルコントロール・アプリケーション(Canopy、Netspark mobile)、ISPソリューション(現在イスラエル国外にも展開)、B2B市場向けソリューション(UGC企業、モデレーション企業など)です。

私たちは、私たちのアルゴリズムをテストすることに決め、その精度を、AWS、Google、Microsoftといったテックジャイアントのアルゴリズムの精度と比較しました。

確かに最初は半信半疑で、地球上で最も優秀なエンジニアにアクセスでき、資金も無限にあるような強力な企業に勝てるわけがないと言われました。それでも私たちは、自分たちをよく観察し、ポルノグラフィーのモデルをテストすることで、自分たちの真価を見極めることにしたのです。

方法論

ポルノコンテンツを分類し、ポルノを100%ブロックすることを考えるならば、単純にすべてのコンテンツをポルノを含むものとして分類することができます。この方法では、ポルノを含むコンテンツはすべて実際にポルノとして分類されますが、ポルノでないコンテンツもすべてポルノを含むものとして分類されます。

ポルノを含むコンテンツだけを検出・分類し、ポルノ以外のコンテンツを無視しようとする、非常に精度が高くないといけないのです。

まず、母集団が比例しないことを理解する必要があります。例えば、1,000枚の画像のうち、ポルノが含まれているのは10枚だけで、90%の確率でポルノと識別するモデル(10%の誤り)と、90%の確率でポルノでない画像を識別するモデル(10%の誤り)を使用すると、統計的には90%で正しく、10%で誤った分類となるわけです。

つまり、上記の例では、1枚のポルノ画像を正しく分類すると、10枚の画像を誤って分類してしまうことになります。つまり、1,000枚の画像のうち、108枚をポルノが含まれる画像として分類することになりますが、これは99枚が間違いで、9枚が正しいということです。

上記の例から、画像解析は妥協の産物であるという単純な事実が明らかになりました。より正確であることを選択し、この場合、アルゴリズムが行う分類はすべて正しいが、アルゴリズムはかなりの数の画像を見逃してしまうか、より正確でないことを選択し、分類のいくつかが間違っていることになります。

したがって、異なるアルゴリズムの精度を比較できるようにするために、すべてのアルゴリズムに同じ精度レベルを選択する必要があります。そこで、アルゴリズムができるだけ多くのポルノ画像／動画を識別できるように、精度レベルを設定しようとしていました。その結果、最も近い競合他社が96.5%の精度であったのに対し、私たちのモデルは99%の精度を達成し、驚きました。そこで、私たちのモデルも同じ96.5%の精度に制限しました。

比較

比較したモデル。

比較を行うために、以下のAPI'を使用しました。

1. Netspark - <https://rapidapi.com/user/Netspark>
2. **Microsoft Azure** - <https://docs.microsoft.com/en-us/azure/cognitive-services/content-moderator/image-moderation-api>
3. **AWS**
<https://docs.aws.amazon.com/rekognition/latest/dg/moderation.html>
4. **Google** - <https://cloud.google.com/video-intelligence/docs/analyze-safesearch>

比較方法。

- **Netspark** - ビデオは、モデルが高い確率で結果を返す場合のみ、ポルノとして分類されました - 「非常に可能性が高い」と「可能性が高い」です。
- **Azure** - Azureのビデオエンジンが見つからなかったため、偏りのない比較を行うのに問題がありました。
フレーム上で動作するエンジンのみ)でフレームモデルをテストしました。
この方法は、Azureにとって最も有益な方法でした(他の方法では、精度がさらに低くなります)。
- **AWS** - AWSのモデルは動画中のヌードフレームの数をカウントしているため、フレームの1%以上がポルノと分類された場合、動画全体をポルノが含まれていると分類することにしたのです。
- **Google** - ビデオは、モデルが高い確率で結果を返す場合のみポルノとして分類された - 「非常に可能性が高い」と「可能性が高い」。

動画比較結果

下の表は、比較したアルゴリズムでポルノと分類された動画の割合を示しています。

カテゴリー	説明	Netspark	Azure	AWS	Google
No-porn 簡単	偽陽性-割合は低い方が良い (0%)	0.50%	6.47%	0.50%	1.00%
No-porn 高難度	偽陽性-割合は低い方が良い (0%)	10.00%	34.00%	31.00%	16.50%
Porn	トゥルーポジティブ - 高ければ 高いほど良い (100%)	96.50%	94.26%	96.50%	94.75%

上の表に明記されている結果を見て、私たちはとても驚きましたが、同時にとても嬉しくもありました。Netsparkのモデルは、全ての大手のモデルよりも著しく正確であることがわかります。

Non-porn-Easyの結果を見ると、NetsparkとAWSは0.5%しかポルノと誤認していないのに対し、Microsoftは6.47%、Googleは1%をポルノを含むコンテンツと誤認しており、同様の結果が出ている。

Non-porn-Difficultの結果を見ると、Netsparkがポルノと誤認したのは10%に過ぎないのに対し、Googleは16.5%(Netsparkの65%増)、MicrosoftとAmazonは30%以上(Netsparkの3倍以上)のコンテンツをポルノと誤認していることが分かる。

Pornカテゴリーの結果を見ると、NetsparkとAmazonだけが96.5%のコンテンツをポルノと認定していることがわかる。

上記のように、Netsparkがポルノの可能性があると分類した結果を考慮した場合、99%のポルノを検出することができ、MicrosoftとGoogleは95%未満しか識別することができません。

フレーム比較結果

私たちが入手したポルノデータベースには、「Porn」カテゴリの下に、ポルノフレームと非ポルノフレームが混在していました。私たちの理解に従ってフレームを分類することもできましたが、そうするとデータベースの信憑性を脅かすことになり、比較の整合性に疑問が生じかねません。そのため、Non-pornカテゴリのみで比較することにしました。

ポルノ以外のビデオにはポルノフレームは含まれず、ポルノビデオにはポルノフレームと非ポルノフレームが混在しているので、誤検出率、つまり、モデルが何回フレームをポルノと誤って分類したかを調べることができます。

下の表は誤検出率、つまりアルゴリズムによればポルノでないビデオにポルノフレームが含まれている割合を示しています。

カテゴリー	説明	Netspark	Azure	AWS	Google
ノンポルノ - 簡単	偽陽性 - 少ない方が良い (0%)	0.00%	6.47%	3.98%	10.45%
ノンポーン 高難易度	偽陽性 - 少ない方が良い (0%)	3.50%	34.00%	22.00%	37.50%

上記の結果を調べたとき、私たちは完全に驚かされました。まさか、競合他社の6倍もの精度を誇るとは.....(笑)。

Googleのアルゴリズムで、同じようにポルノの可能性があると分類された結果も考慮すると、Googleのアルゴリズムではポルノ検出の精度は95.5%となります。ポルノ以外の簡単なカテゴリでは1.5%、ポルノ以外の難しいカテゴリでは33%の誤検出率となります。

結論

私たちのモデルは、競合他社よりもはるかに精度が高いのです。

当社のモデルは、最高レベルの真陽性結果を出すだけでなく、偽陽性分類も競合より大幅に低く、難しいフレームでは競合の6倍以上の精度となっています。

詳細については、sales@netspark.com までお問い合わせください。

[1] Sandra Avila, Nicolas Thome, Matthieu Cord, Eduardo Valle, Arnaldo de A. Araújo.Araújo.”画像表現におけるプーリング: Visual Codeword Point of View”.Computer Vision and Image Understanding, volume 117, issue 5, p. 453-465, 2013.DOI: <http://dx.doi.org/10.1016/j.cviu.2012.09.007>