

「AI（考える）」技術

AIによる行動認識技術

「行動認識」って言葉を最近耳にします。その「行動認識」がどんな技術で、どんな風に私たちの生活の中で役立つのか紹介します。

最終更新日 2021年11月26日

もくじ

- ↓ 行動認識って？
- ↓ AIによる行動認識は必要なの？
- ↓ 従来技術－骨格に基づく行動認識技術
 - 【Step1】画像に写る人の関節の位置を学習し、骨格を推定する
 - 【Step2】推定した骨格から人の行動を認識する
- ↓ 富士通の開発技術（Actlyzer アクトライザー）
 - 【Tech1】短時間で現場に導入可能！（製品化済）
 - 【Tech2】正解率を上げて、Actlyzerの応用範囲を広げる技術！（開発中）
- ↓ 富士通のAIによる行動認識技術の応用例（防犯カメラ以外）
- ↓ 小話
- ↓ 関連リンク

行動認識って？



突然だけど、この女の子は、どんな動作をしているか当ててみて？



簡単じゃん！1番は朝の挨拶、2番はお客様をご案内、3番はバイバイしているでしょ。



はい、正解！

3人とも同じように右手をあげていても、人は音声や細かい動きを見て、どんな行動なのかわかりますよね。それは人の成長過程で、どんな行動がどういう意味を持つかなどを少しずつ教えてもらったり、覚えたからわかるんですよね。



確かにそうですよね。知らない行動だったら、私たちだってどう答えて良いかわかりませんね。

同じことをAIでやろうとすると、前もって大量のデータ（教師データ）を準備して、それを正しく判断できるようにAIが学習する必要があります。



人は簡単にできることだけど、AIには大変そうですね

AIによる行動認識は必要なの？



AIによる行動認識が難しいなら、人による（目視）行動認識だけでいいんじゃない？ AIが行動認識する必要ってあるかな？



防犯カメラなどは、人間の代わりにAIが人の行動をみってくれるって何かで読んだわ。

はい、そのとおりです。

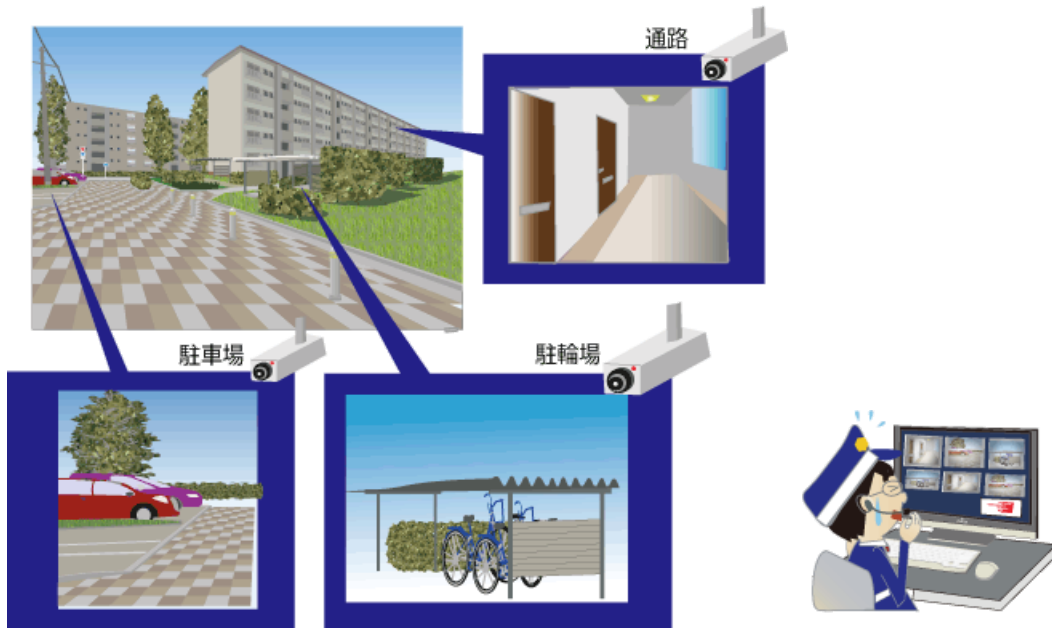
防犯カメラの数が多いと映像量が膨大になります。人は疲れたり目が順応してしまい、見逃してしまうこともあるかもしれません。それに比べてAIは疲れを知りません。必要であればコンピューターを増やして並列（同時に複数のPCで実行させること）処理することもできます。AIによる行動認識を取り入れると、膨大な映像から不審と思われる箇所だけをピックアップすることができ、最終確認をする人の負担を減らす効果を期待できます。また、何かが起こる前触れの検知もできます。



最近では、防犯カメラはビルや工場だけでなく、マンションや一軒家にも設置されているよね！うちの玄関にもあるよ！



確かに膨大な映像量になりますね。AIによる行動認識ができれば、人の作業負担を減らすことができますから、助かりますね！



従来技術一骨格に基づく行動認識技術

現在一般的に使われている行動認識技術を紹介します。この技術は、2つのStepで成り立っています。

- ・【Step1】 画像に写る人の関節の位置を学習し、骨格を推定する
- ・【Step2】 推定した骨格から人の行動を認識する

【Step1】 画像に写る人の関節の位置を学習し、骨格を推定する

【Step1】は、**教師データ**を使って**ニューラルネットワーク**に骨格を学習させ、人の骨格を推定できるようにする技術です。



その「**教師データ**」って何ですか？

AIが学習するための教科書のようなものです。ここでは、
「**人間が映っている動画や写真のファイル**」と
「**その動画や写真の中の人の関節の位置情報を記録したファイル**」を用意します。

（関節の位置情報というのは、人体の全ての関節ではありません。行動認識に必要な「肩」「ひざ」「手首」「腰」など全部で数十点です。各写真の中の人の関節がどこにあるのか、人間が位置を示した数値が記録されているファイルです）



その教師データは、何枚くらい必要ですか？10枚？100枚？

いえいえ、**数十万～数百万枚**です。



エー!! 教師データを用意するだけで大変な作業だし、時間がかかったらうわ！



数十万～数百万枚の写真や動画を用意 1枚ずつ関節部分に印をつけて、位置情報ファイルを用意



もうひとつ教えて欲しい言葉があって、「教師データを使ってニューラルネットワークに骨格を学習させる」の説明で使われている「**ニューラルネットワーク**」って何ですか？

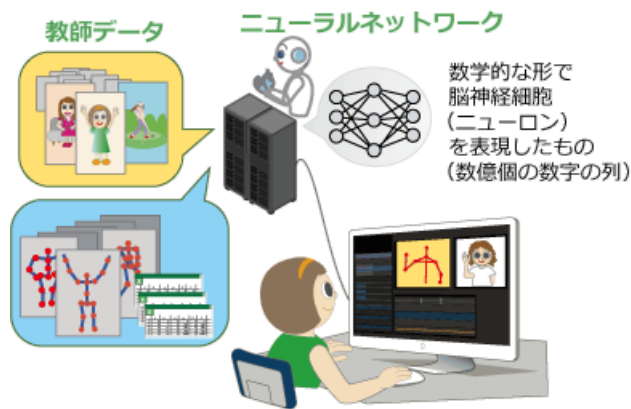
AIはコンピューターを中心とする人工的なシステムです。ここではディープラーニングといって、自動で機械学習する方法を使用しています。そのディープラーニングで使っている、人間の神経細胞をモデルにした複雑な計算をおこなう部分を「ディープニューラルネットワーク」あるいは単に「ニューラルネットワーク」といいます。



なんだかすごいシステムですね。そのニューラルネットワークに、どうやって骨格を学習させるのですか？

はい、教師データの「写真や映像」と、「関節の位置を書いたファイル」を使って、人の関節の位置を学習させます。例えば「写真」の中に人物を見つけて、その人物の関節の位置を示せるようになるまで、繰り返し学習します。





ディープラーニング

大量の教師データを使って、人の関節の位置を学習。
人の関節の位置を示せるようになるまで繰り返す。

骨格を推定



関節の位置を順番に
線でつなげて
骨格を推定します!



関節の位置を推定できれば、人の骨格がわかるってことですね。

はい、その通りです。関節の位置と骨格は連動していますから、関節を
順番につなげて人の骨格を推定します。



【Step2】推定した骨格から人の行動を認識する

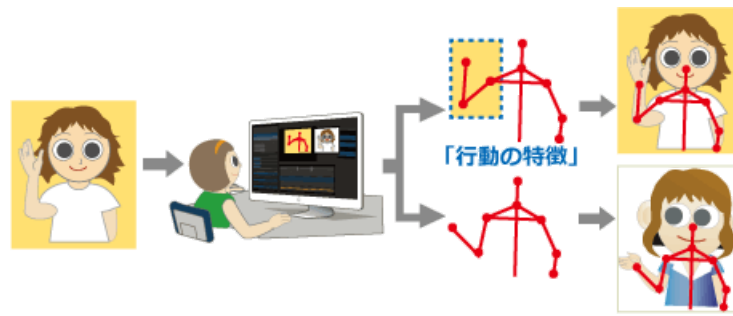
【Step2】は、推定した骨格から「**行動の特徴**」を抽出し、それがどう
いう行動なのかを判断する技術です。



「**行動の特徴**」って何ですか?

はい、**人の行動を判断する上で大事な特徴的な部分**のことです。例え
ば、手の曲がり具合が異なるなら、手の部分がその写真の「**行動の特
徴**」となり、どういう行動なのか推定します。





同じように右手をあげていても、「おはようの挨拶」か、「こちらへどうぞと案内している」かってことですね。しかし「行動の特徴」が少ない場合、判断に困ることもありそうですね。

はい、良いところに気がつきましたね。それでは、従来技術の課題を整理してみましょう。



— 従来技術の課題 —

- ・ 大量の教師データを準備するので、実際に使えるようになるまでに時間がかかる。
- ・ 「行動の特徴」の変化が少ないと、AIが間違っただけ判断することもある。

これらの課題を解決する富士通の開発技術を紹介します。



富士通の開発技術（Actlyzer アクトライザー）

従来技術の課題を解決する2つの開発技術を紹介します。

【Tech1】 短時間で現場に導入可能！

約100種類の基本動作を独自に定義して、その動作をすでに学習済みにした技術です。

従来技術の課題1つ目は、お客様が行動認識技術を使いたいと思って導入するまでに時間がかかっていたことです。



従来技術では、たくさんのデータを用意して学習させる必要がありましたね。もっと短時間で導入できるといいですね。

はい、富士通で開発した「**Actlyzer**」は、人の行動を構成する約100種類の基本動作を独自に定義して、すでに学習済み（ディープラーニング）です。その基本動作を組み合わせることで、認識した行動を指定できます。



えー！！基本動作がすでに学習済みなら、教師データを準備したり学習させる必要が無いってことですよね！すごすぎ…。



Actlyzerの約100種類の基本動作って、どんな動作なのかな。教えてください。

私たちが日常でよく使う動作で、座る、走る、立つ、歩く、起きるなどです。



**約 100 種類の「人の基本動作」を独自に定義
(例)**

01: 立っている	02: 起きる	03: 走る	04: しゃがむ
07: 歩く	08: 倒れる	09: 飛ぶ	10: はう
26: 顔・上をむく	27: 顔・下をむく	28: 顔・左をむく	29: 顔・右をむく
50: 右足・上	51: 右足・下	52: 右足・左	53: 右足・右



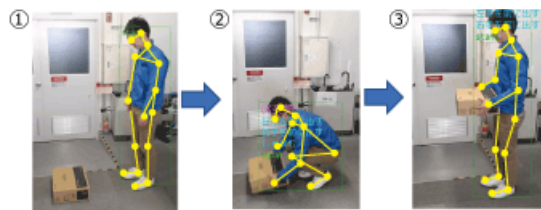
約100種類の基本動作を使って、Actlyzerはどうやって人の行動を認識しているのですか？

まず、

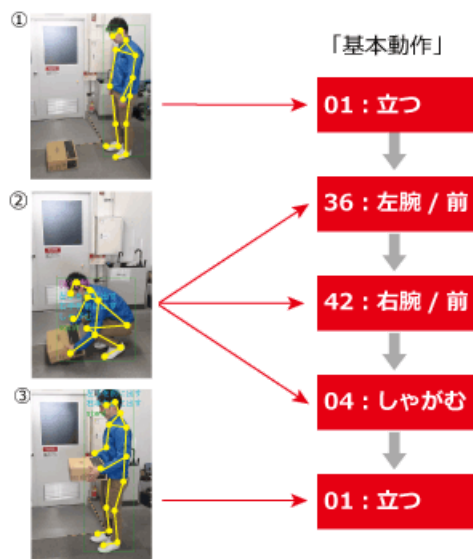
- ・対象となる人の骨格を推定、
- ・その骨格から「基本動作」のどれに相当するのか推定、
- ・その「基本動作」を登場順に並べてその人の行動を認識しています。



例えば、床の上の荷物をも持ち上げる動作をしている人がいます。まずは骨格を推定します。



その人の骨格から、「立つ」、「左腕前」、「右腕前」、「しゃがむ」、「立つ」という基本動作を推定しました。そして、順番にそって組み合わせます。



これらの基本動作の組み合わせと順番を別途定義済みの行動データベースと照らし合わせて、この人は「荷物を持ち上げる」という動作をActlyzerは判断します。（実際の現場では、基本動作の組み合わせと順番と、「行動の発生場所」「対象人物の属性」などの情報も含めることで、さらに正確に判断しています）



たくさんの教師データを用意したり、学習させる必要が無く、基本動作の組み合わせと順番を記述するだけで導入可能だから、短時間なんです。納得です！

【Tech2】正解率を上げて、Actlyzerの応用範囲を広げる技術（開発中）

■正解率を上げる技術



正解率を上げるために、どんなことをしているのですか？

従来技術Step2の「推定した骨格から人の行動を認識する」の説明で「行動の特徴」の抽出について話しましたね。



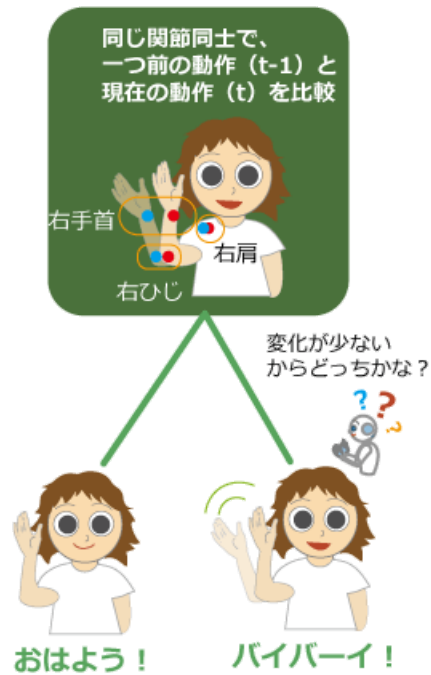
はい、「人の行動を判断する上で、大事な特徴的な部分」と教わりました。

その「行動の特徴」の情報量が少ないと、AIは判断に迷うことがあります。



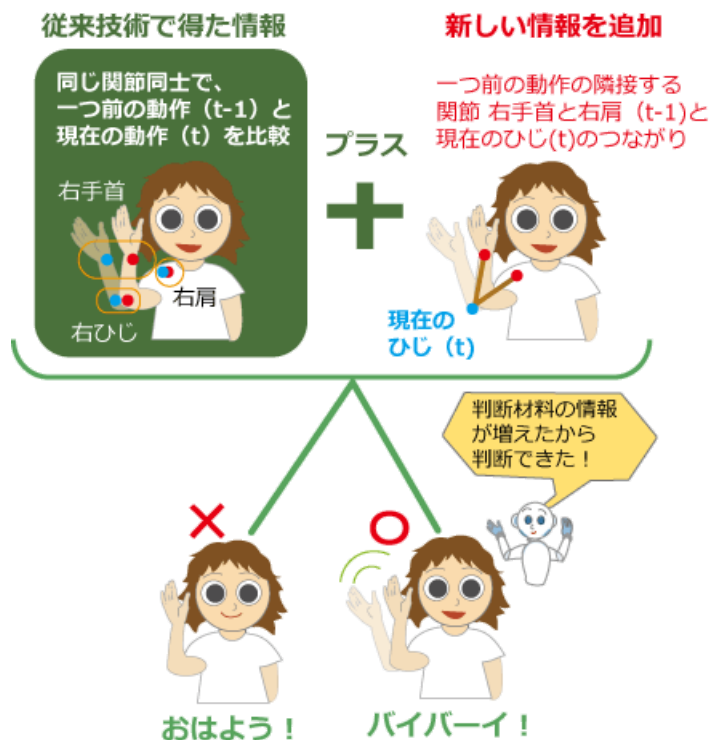
（従来技術）

人の行動を判断するために、一つ前の動作の関節の位置と同じ関節の現在の位置を比較していました。一つ前の動作と現在の動作の変化が少ないと迷うことがあります。



（開発技術）

同じ関節同士の比較に加えて、一つ前の動作の隣接する関節についても「行動の特徴」として抽出します。



開発技術は、同じ関節同士の一つ前の動作時と現在の位置の比較に加えて、一つ前の動作の隣接する関節についても「行動の特徴」として抽出する、と書いてあります。「一つ前の動作の隣接する関節とつなぐ」とはどういうことですか？

まず、隣接する関節ですが、「右ひじ」に注目した場合、隣接するのは「右手首」と「右肩」になりますよね。



はい、わかります。
それで、なぜ一つ前の動作の隣接する関節とつなぐのですか？

現在の動作の関節同士はすでにつないでいます。その他に「右ひじ」にとって一番影響のあるのは、「一つ前の動作の右手首と右肩」になるからです。



それでは、どうして「首」や「左手首」はつながないのですか？

たとえば、「右ひじ」から離れた位置にある「首」の関節や「左手首」の関節は、「右ひじ」が動いた影響で、連動して動くということはありませんね。連動する関節の位置情報は必要ですが、「右ひじ」からみて「首」や「左手首」のように、連動しない関節の位置情報は不要になります。



右手をあげて「バイバイ」動作の際に、「右ひじ」と「首」「左手首」は連動しないので、AIにとって不要な情報になります。



でも、AIって情報量が多い方が正解率が上がるなら、体全部の関節とつないだ情報をAIに与える方が良いのではないですか？

AIにとって情報量が多いのは良いですが、多すぎても間違ってしまう原因になることがわかりました。例えば、右手で「バイバイ」の動作の際、「左手をあげている左手首の情報」までAIにあたえてしまうと、「左手をあげていなければ、(右手の) バイバイではない」と判断するかもしれません。そのため、今回は「連動する関節」というのを隣の関節までに絞り込んでいます。



なるほど～。AIが判断するために、**ちょうど良い情報量**っていうのがあるんですね。

はい。今回の開発技術で、400の行動パターンから正しい行動をどれくらい認識できるかという実証実験で、**世界一の認識精度を達成し、世界トップクラスの行動認識の研究者が集う学会で、その成果が認められました。**



すごい、ヤッタネ！



専門家が集う学会で、
研究成果が認められました！

■Actlyzerの応用範囲を広げる技術



どうして、正解率が上がると、Actlyzerの応用範囲が広がるのですか？

正解した行動を常にActlyzerの基本動作に取り込んでいきます。結果的に、どんどん学習済の基本動作が増えていきます。



「Actlyzer」ですでに学習済みの基本動作
(例)

01: 立っている	02: 起きる	03: 走る
07: 歩く	08: 倒れる	09: 飛ぶ
26: 腕・上をむく	27: 腕・下をむく	28: 腕・左をむく
50: 右足・上	51: 右足・下	52: 右足・左

基本動作が増える⇒Actlyzerの応用範囲が増える



ニューラル
ネットワーク

数学的な形で
脳神経細胞（ニューロン）
を表現したもの
(数億個の数字の列)



ん？？学習済の基本動作が増えると何が良くなるのですか？

例えば、今までの基本動作の中には、「挨拶のハグ」や「ケンカでつかみかかっている」などの複雑な動作は入っていません。それらを認識しようとする、新たにたくさんのデータが必要となります。Actlyzerは一度認識すれば、それが基本動作として蓄積されていきます。





どんどん進化していくってことですね！

その通りです。Actlyzerの特徴である「基本動作」の数が増えるほど、その組み合わせや順番を記述するだけで行動を認識できる応用範囲が広がります。



すごいですね！

富士通のAIによる行動認識技術の応用例（防犯カメラ以外）

複雑な行動も手軽に認識可能！

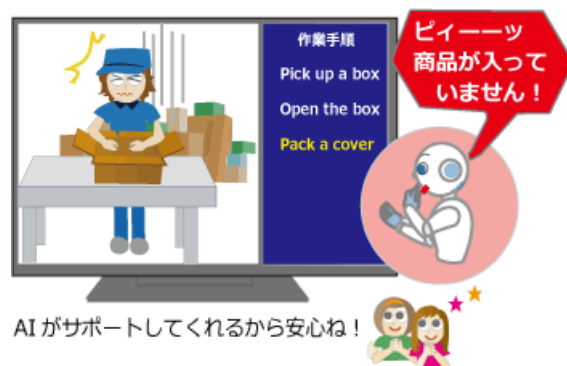


基本動作の組み合わせや順番を記述するだけで、特定の行動を認識できます！



「スマホを見ながら自転車に乗っている人認識したい」という複雑な行動の要望があった時に、富士通の「Actlyzer」は基本動作の組み合わせや順番を記述するだけで、特定の行動を判断できます。

作業のうっかりミスの防止に！



製造現場で同じ作業をして、疲れてくると「うっかり」なんてこともあるかもしれませんね。

無理な姿勢での事故防止に！



無理な姿勢で重い荷物を持ち上げようとするとうちに負担がかかりますね。ぎっくり腰などの事故防止に役立ちますね。

大事故の防止に！



駅のホームでスマホをみながら歩いている人が、線路へ落ちてしまうなどのニュースを聞いたことがありますね。

体調不良者の早期発見に！



体調不良で自分から声をだすことができない時ってありますよね。そんな時に見つけてくれると助かりますね。

小話

「Actlyzer」という名前が決まるまでの物語

行動分析技術の研究開発しているけれど、社内でさえ、よく理解されているとは言えない技術・・・なぜ広まらない？？さらに、研究技術に名前がないと、資料が書きにくくて困る・・・

それでは、名前を付けよう！ということになった。

技術をアピールできる名前で、覚えやすく、カッコいいのがいい！でも、すでに他で使われている名前は、商標登録できないから調べる必要がある。（登録まで1年近くかかるらしいから、早めに決める必要がある。時間がない！）まずどうしようか、名前を決める進行係がてっとりばやく「部内で募集してみよう！」と実際に募集してみた。

考えが甘かった・・・。

部員はどんなネーミングが良いのかわからず、候補を上げたくてもあげられない。つまり候補数が集まらない。困った進行係は、技術を開発するように整理して考えてみることにした。つまり、ネーミング方針を決めて、少人数のブレインストーミングをする。候補をだしやすい雰囲気を作る。これでやってみよう！

実際にやってみた。

さすが考えるのが仕事の研究員、でるわでるわで、300弱の候補が上がった！

ここからは進行係のウデのみせどころ。他ですでに使われているものや、名前から技術を連想できないものは脱落。イメージするものが似ているもの同士で分類。残ったもので、部内で電子投票をした。結果が割れるかと思ったら、「Actlyzer」に投票した人が多数のため決定！

「Actlyzer」は、action（行動）と analyzer（分析装置、分析者）の2つの単語から作った造語です。

（進行をまかされていた研究員さん、お疲れ様でした！ By 講座筆者）



関連リンク

プレスリリース

- ＞ 世界最高精度を達成！人の複雑な行動を、連動する各関節の位置や接続関係から高精度に認識できるAI技術を開発（2021年1月13日）
- ＞ 正しい手洗い動作を判定する映像認識AI技術を開発（2020年5月26日）
- ＞ 映像から人の様々な行動を認識するAI技術「行動分析技術 Actlyzer」を開発（2019年11月25日）