

## Attribute Based Learning

### Introduction

เป็นการกำหนด attributes ให้กับรูปต่าง ๆ เพื่อนำมาแบ่งแยกวัตถุในภาพว่าเป็นวัตถุอะไร โดยกำหนด attributes จากลักษณะภายนอกที่สามารถมองเห็นได้เช่น จะแยกรูปผลไม้จากสี ลายแพทเทินของผลไม้ หรือ จะแยกลักษณะท่าทางของคนก็แยกจากกริยาท่าทาง วัตถุ สภาพแวดล้อมเป็นต้น ทั้งนี้การกำหนด attribute ต่อหนึ่งรูปให้เยอะจะเป็นการเพิ่มความแม่นยำของการบอกว่าภาพดังกล่าวเป็นภาพอะไร และ เพิ่มความหลากหลายของภาพที่สามารถบอกได้จาก attributes

ซึ่งการมองและแยกแยะวัตถุเป้าหมายเป็น attributes นอกจากจะทำให้เราสามารถแบ่งแยกเป้าหมายที่เราเรียนรู้มาก่อนหน้าแล้วยังสามารถแยกแยะเป้าหมายใหม่ที่ไม่เคยเห็นมาก่อนด้วยการเพิ่ม attributes ใหม่ในการแยกแยะได้ กระนั้นวิธีนี้ต้องคิดถึงวิธีการว่าจะกำหนด attributes อย่างไร จะกำหนดเองด้วยสายตาแล้วกำหนด attributes หรือ จะให้คอมพิวเตอร์เรียนรู้จากคำบรรยาย แล้วจะเลือกรูปให้คำบรรยายอย่างไรจึงจะมี attributes ที่เหมาะสมครบถ้วน

กล่าวคือจะค่อย ๆ ตัดสินใจลงไปจาก attributes ที่ภาพดังกล่าวมีเป็น Decision Tree เพื่อตัดใจว่าภาพดังกล่าวเป็นภาพอะไร




### How it works

เราจะเทรน features ของภาพต่าง ๆ เป็นส่วนย่อย ๆ ก่อนเช่นเราเทรนเพื่อหาว่า feature นี้คือล้อรถ นี้คือลายทาง นี้คือทรงกลม นี้คือสีแดง เป็นต้นแล้วนำ features ย่อย ๆ เหล่านี้ไปบอกว่ามันเป็น attributes ไດ หรือ เราจะเทรนด้วยการค่อย ๆ บอกว่านี่คือ

attributes ใดในภาพด้วยมือ หรือ สายตาก็ได้ทั้งนี้ทั้งสองวิธีมีข้อดี ข้อเสียต่างกันหรือ อาจจะนำมาใช้ร่วมกันได้เพราะการเทรนจะสามารถทำให้เรารู้ถึง features ที่ซ่อนอยู่ในภาพได้ ในขณะที่เดียวกันการกำหนดด้วยสายตามนุษย์จะทำให้สามารถกำหนด attributes ได้เฉพาะเจาะจงมากขึ้น













ภาพ1.1 ทางซ้ายมือแสดงถึงภาพที่สามารถบรรยายเป็น attributes ได้ด้วยภาพทั้งภาพ ทางขวาแสดงถึงภาพที่บรรยายเป็น attributes เป็นส่วน ๆ (segments )

<b>green</b>	salad (.84), green lizard (.73), bonsai (.52), pesto (.43), saute (.37), daisy (.30) pot-au-feu (.12), salsa (.12), roughage (.11), cow (.11) 
<b>white</b>	kuvasz (.70), Saint Bernard (.67), clumber (.65), wirehair (.62), foxhound (.60) sheet (.49), gerbil (.48), Persian cat (.48), sail (.45), bullterrier (.43) 
<b>round</b>	egg yolk (.75), basketball (.68), button (.63), goulash (.56), basket (.49), ramekin (.47), ball (.42), pot (.42), veloute (.39), miso (.37) 
<b>long</b>	kirsch (.83), sail (.77), rorqual (.74), police van (.72), fork (.69), rack (.67), killer whale (.58), window (.54), transporter (.50), pool table (.49) 
<b>striped</b>	barn spider (.36), daisy (.17), zebra (.17), echidna (.16), backboard (.13), drum (.12), coloring (.12), roller coaster (.12), bridge (.11), colobus (.11) 
<b>wet</b>	rorqual (.59), sidecar (.55), orangeade (.53), flan (.52), screwdriver (.47), killer whale (.44), bowhead (.43), maraschino (.41), dugong (.40), porpoise (.40) 

ภาพ1.2 แสดงตัวอย่าง attributes ที่เรียนของภาพพร้อมคะแนน Median จาก classifier

เมื่อเราสามารถกำหนด attributes ต่าง ๆ แยกย่อยออกมาเป็น features ได้แล้วก็นำ attributes ต่าง ๆ เหล่านี้ไปกำหนดภาพต่าง ๆ อีกทีว่าภาพนั้นควรจะ label ว่าเป็นอะไรเมื่อมี attributes ต่าง ๆ เหล่านั้นเช่นควรจะเป็นลูกบาสเมื่อเป็นทรงกลม สีส้ม มีลายตัด หรือ ควรจะเป็นรถยนต์เมื่อมีล้อ มีความสะท้อนแสง มันเงา เป็นต้น ทั้งนี้ขึ้นอยู่กับ การกำหนด attributes ของภาพ และ ข้อมูลว่าอะไรเป็นอะไร ถ้าภาพดังกล่าวพึ่งพา attributes ที่มีความ Generalize มากเกินไปจะทำให้บอกได้ยากกว่าภาพดังกล่าวเป็นภาพอะไร

						
<b>chestnut:</b> brown,smooth		<b>0.52</b>	0.16	0.12	0.12	0.08
<b>green lizard:</b> green, long		0	<b>0.84</b>	0	0.12	0.04
<b>honey badger:</b> black, gray, rough, furry		0.32	0	<b>0.60</b>	0.04	0.04
<b>zebra:</b> black, white, striped, smooth		0.36	0.08	0.40	<b>0.08</b>	0.08
<b>spitz:</b> white, furry		0.08	0	0.36	0.08	<b>0.48</b>

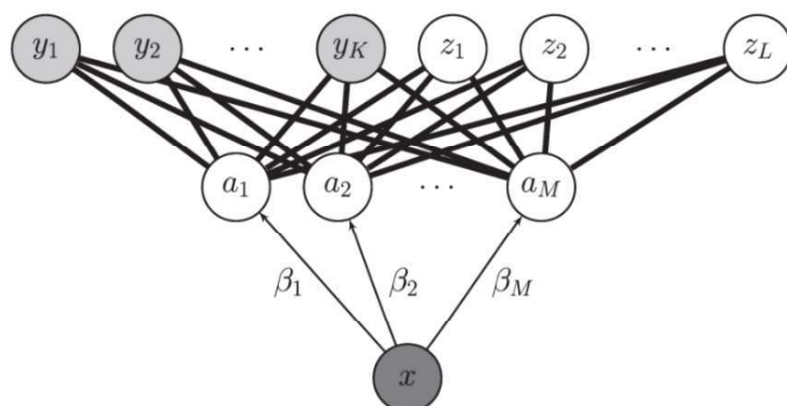
ภาพ 1.3 แสดงตัวอย่างผลลัพธ์การเลือกภาพพร้อมคะแนน

จากตัวอย่างภาพ 1.3 จะเห็นว่าเลือกม้าลายผิด เนื่องจากม้าลายฟังการบ่งบอกจาก attributes "Striped" มากกว่ามันคือม้าลาย แต่ว่าการหา features "Striped" ทำได้ไม่ดีพอ ทำให้ความสามารถในการม้าลายแย่ลง

ประเภทของ Attribute based learning

Direct attribute based learning

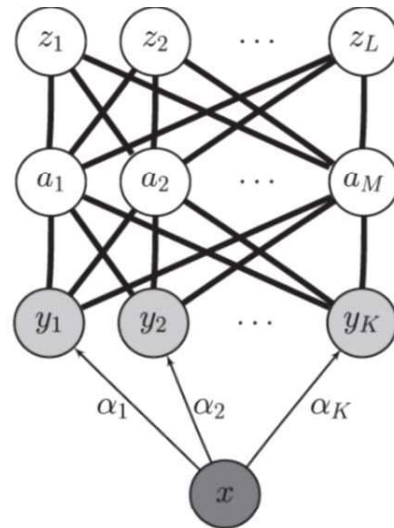
ตรงตัวคือการเทรน attribute เพื่อนำไปแบ่งแยกประเภทของ object โดยตรง โดยการเทรน attribute แล้วนำไปหา probability ของ attribute เหล่านั้นแล้วคัดกรอง attribute ของ object



(b) Direct attribute prediction (DAP)

## Indirect attribute based learning

เป็นการเทรน class อื่น ๆ ก่อนหน้าเพื่อหา attribute ที่อยู่ในคลาสเหล่านั้นแล้วนำมาใช้แบ่งแยกคลาสอื่น ๆ ที่จะเข้ามาใหม่เพิ่มเติม



(c) Indirect attribute prediction (IAP)

คะแนน กับ การเลือกจาก Probability

การนำคะแนนมาใช้งานนั้นคิดจากสมการ

Semantic attributes probability

$$p(z|x) = \sum_{a \in \{0,1\}^M} p(z|a)p(a|x) = \frac{p(z)}{p(a^z)} \prod_{m=1}^M p(a_m^z|x)$$

Class probability

$$c(x) = \arg \max_{l=1,\dots,L} \prod_{m=1}^M \frac{p(a_m^{z_l}|x)}{p(a_m^{z_l})}$$

เช่นถ้ามี 1 attribute สำหรับคลาส z ( เมื่อ x เป็นรูปที่นำเข้ามา และ a คือ attribute )

$$P(Z|X) = P(Z|0)P(0|X) + P(Z|1)P(1|X)$$

ทั้งนี้ จะเห็นว่าในกรณีนี้ attribute เป็นได้ไม่ 0 ก็ 1 เพื่อบ่งบอกสถานะของ attribute ดังนั้น probability ของกรณีนี้คือ  $P(Z|X) = P(Z|1)P(1|X)$  ซึ่งถ้าหากมี attribute ที่มากขึ้นก็จะเพิ่มเข้าไป

กรณีตัวอย่างมี 2 attribute และรูปดังกล่าวจะเป็น class Y1 ถ้ามี attribute ทั้งสองตัวเป็น 1

$$P(Z|(0,0))P((0,0)|X) + P(Z|(0,1))P((0,1)|X) + P(Z|(1,0))P((1,0)|X) + P(Z|(1,1))P((1,1)|X)$$

ในกรณีนี้ค่าของ probability คือ  $P(Z|(1,1))P((1,1)|X)$

## Conclusion

เทคนิควิธีการ Attributes learning เป็นการมองภาพว่าในภาพมีอะไรบ้างมาประกอบกันจนบ่งบอกความหมายของภาพนั้นได้ ทั้งนี้การใช้ทั้งตัวมนุษย์ร่วมกับคอมพิวเตอร์ในการกำหนด attributes จะช่วยเพิ่มความสามารถให้วิธีดังกล่าวอย่างเห็นผลได้ นอกจากนี้วิธีดังกล่าวยังเป็นการแสดงแนวคิดใหม่ในการวิเคราะห์ภาพ แทนที่จากเดิมเราให้ภาพทั้งภาพในการวิเคราะห์เราก็ให้ภาพเป็นชิ้นส่วนไป เช่นเราให้หน้าต่าง ล้อรถ ไฟหน้า กับรถยนต์ เป็นต้น

## Reference

[http://cvgl.stanford.edu/papers/cvpr11\\_liu\\_a.pdf](http://cvgl.stanford.edu/papers/cvpr11_liu_a.pdf)

<https://papers.nips.cc/paper/3217-learning-visual-attributes.pdf>

<http://ai.stanford.edu/~olga/papers/eccv10workshop-Attributes.pdf>