

Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer

- Neste artigo é estudado o problema de classificação de objetos quando as classes de treino e as classes de teste são disjuntas (ZSL);
- Para lidar com este problema, é introduzida a **classificação baseada em atributos** (*attribute-based classification*). Este tipo de classificação permite a deteção de objetos **baseada na descrição** de alto nível especificada por humanos **dos objetos que se pretendem classificar**, em vez de imagens de treino. Esta descrição consiste em atributos semânticos arbitrários, como por exemplo a forma, a cor ou mesmo informação geográfica;
- Algumas propriedades podem ser pré-aprendidas a partir de *datasets* de imagens que não estejam relacionados com a tarefa em questão, o que permite que as novas classes possam ser detetadas com base na sua representação de atributos, sem necessidade de uma nova fase de treino;
- É neste artigo que é apresentado o *dataset Animals with Attributes* (AwA);

Introdução

- Os métodos para reconhecer objetos, seguindo a abordagem típica, necessitam de muitos exemplos de treino devidamente classificados para que consigam atingir um bom valor de *accuracy* na classificação dos objetos;
- A ideia é o desenvolvimento de um sistema que consiga detetar objetos a partir de uma lista de atributos de alto nível. Os atributos servem como que uma camada intermédia num classificador e possibilitam que o sistema detete classes de objetos para os quais não existe nenhum exemplo de treino;
- Para aprender os atributos podemos usar as imagens de treino misturando imagens de várias classes de objetos. Para aprender o atributo "ricas" podemos usar imagens de zebras; para aprender o atributo "amarelo" podemos usar tigres, juntamente com canários, etc;
- É possível obter conhecimento sobre atributos usando diferentes classes de objetos e vice-versa, cada atributo pode ser usado para a deteção de muitas classes de objetos.

Transferência de informação através da partilha de atributos

- O uso de atributos permite a transferência de informação entre classes de objetos;

Aprender com classes de treino e classes de teste disjuntas

- A tarefa é aprender um classificador para um conjunto de classes que é disjunto do conjunto de classes que foi treinado;
- Os classificadores típicos aprendem um vetor de parâmetros para cada classe de treino. Neste caso, como as classes do conjunto de teste não estão presentes na fase de treino, é impossível aprender um vetor de parâmetros para essas classes;
- De maneira a fazer as previsões sobre essas classes, as quais não dispõem de dados de treino, é necessário introduzir um acoplamento entre as classes Y (*seen classes*) e Z (*unseen classes*). Dado que não existem dados de treino para as classes desconhecidas, este acoplamento não pode ser aprendido a partir de exemplos, mas pode ser introduzido no sistema pelo esforço humano. Ora, isto leva a dois entraves: 1) a quantidade de esforço humano para especificar as novas classes deve ser mínimo, porque a recolha e a classificação de exemplos de treino será uma solução simples; 2) acoplar dados que requerem apenas conhecimento comum é preferível a acoplar dados que requerem conhecimento especializado, porque mais tarde é difícil de obter.

Classificação Baseada em Atributos

- Os objetivos são conseguidos com a **introdução de um conjunto pequeno de atributos semânticos de alto nível *per-class***. Estes atributos podem ser a cor e a forma, para objetos, ou o habitat natural para os animais;
- Se para cada classe $z \in Z$ e $y \in Y$ estiver disponível uma representação de atributos $a \in A$, então é possível aprender um classificador $\alpha : X \rightarrow Z$ transferindo informação entre Y e Z através de A .
- A classificação baseada em atributos é uma das soluções para o problema de aprendizagem com classes de treino e teste disjuntas;
- São introduzidos e comparados o ***Direct attribute prediction* (DAP)** e o ***Indirect attribute prediction* (IAP)**, que são dois métodos genéricos para integrar atributos na classificação multi-classe.
- O **DAP** usa uma camada intermédia de variáveis de atributos para separar as imagens da camada das classes. Durante o treino, a classe devolvida de cada exemplo induz uma etiquetagem determinística da camada de atributos. Consequentemente, qualquer método de aprendizagem supervisionada pode ser usado para aprender os parâmetros *per-attribute* β_m . Na fase de teste, estes parâmetros permitem a previsão dos valores de atributos para cada exemplo de teste, a partir dos quais a nova classe é inferida.
- O **IAP** também usa os atributos para transferir conhecimento entre as classes, mas os atributos formam uma camada que é inserida entre duas camadas de classes (a camada das classes de treino e a camada das classes de teste). A fase de treino do IAP é a classificação multi-classe

comum, onde são aprendidos vários parâmetros α_k . Na fase de teste, as previsões para todas as classes de treino induzem a etiquetagem da camada de atributos a partir da qual uma etiquetagem sobre as classes de teste pode ser inferida.

- A principal diferença entre estas duas abordagens reside na relação entre as classes de treino e as classes de teste. A aprendizagem direta dos atributos (DAP) resulta numa rede onde todas as classes são tratadas de igual forma. Quando as classes de teste são inferidas, a decisão para todas as classes é apenas baseada na camada de atributos. Por outro lado, se a previsão dos atributos for feita indiretamente (IAP), as classes de treino funcionam aqui também como que um intermediário na fase de teste. A derivação da camada de atributos a partir da camada das classes de treino irá atuar como que uma etapa de regularização, que cria combinações sensíveis de atributos e o sistema ficará mais robusto.

Implementação

- Ambos os métodos de classificação em cascata, **DAP e IAP**, podem ser **implementados combinando um classificador** ou um regressor supervisionado para a predição da *image-attribute* ou *image-class* **com um método^[1] de inferência livre de parâmetros** para canalizar a informação através da camada de atributos. São usados modelos probabilísticos.
- Para simplificar, todos os atributos têm valores binários, tal que a representação de atributos $a^y = (a_1^y, \dots, a_m^y)$ para qualquer classe de treino y é um vetor binário de tamanho fixo. Os atributos contínuos podem ser tratados da mesma maneira usando a regressão em vez de classificação.
- Para o DAP, **começamos por aprender classificadores probabilísticos^[2] para cada atributo** a_m , usando todas as imagens de todas as classes de treino como exemplos. A um exemplo da classe y é atribuída a etiqueta binária a_m^y . Os classificadores treinados fornecem as estimativas de $p(a_m|x)$, a partir das quais formamos um modelo para a camada completa *image-attribute* como $p(a|x) = \prod_{m=1}^M p(a_m|x)$. Na fase de teste, assumimos que cada classe z causa o seu vetor de atributos a^z de uma maneira determinística, i.e., $p(a|z) = \mathbb{I}[a = a^z]$, fazendo uso da notação dos parêntesis de Iverson: $\mathbb{I}[P] = \begin{cases} 1 & \text{if } P \text{ is true,} \\ 0 & \text{otherwise} \end{cases}$. Aplicando a regra de Bayes, obtém-se $p(z|a) = \frac{p(z)}{p(a^z)} \mathbb{I}[a = a^z]$ como a representação da camada *attribute-class*. Combinando as duas camadas, podemos calcular a classe de teste z dada uma imagem x :

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

- Na falta de mais conhecimento, assumimos classes de teste anteriores idênticas, o que permite ignorar o fator $p(z)$. Para o fator $p(a)$, assumimos uma distribuição fatorial $p(a) = \prod_{m=1}^M p(a_m)$, usando as médias empíricas $p(a_m) = \frac{1}{k} \sum_{k=1}^K a_m^{y_k}$ sobre as classes de treino

como atributos antecedentes. Mas na verdade assumindo $p(a_m) = \frac{1}{2}$ conseguem-se resultados comparáveis.

- Como regra de decisão $f : X \rightarrow Z$ que atribui a melhor classe de saída de todas as classes de teste z_1, \dots, z_L a um exemplo de teste x , foi usada a predição MAP:

$$f(x) = \arg \max_{l=1, \dots, L} \prod_{m=1}^M \frac{p(a_m^{z_l} | x)}{p(a_m^{z_l})} \quad (2)$$

, onde $p(a_m^{z_l} | x)$ é a probabilidade do atributo dada a imagem x e $p(a_m^{z_l})$ é o atributo prévio estimado pela média dos atributos sobre as classes de treino.

- Para implementar o IAP, apenas é necessário modificar a etapa *image-attribute*: o primeiro passo é aprender um classificador probabilístico multi-classe estimando $p(y_k | x)$ para todas as classes de treino y_1, \dots, y_k . Assumindo uma dependência determinística entre classes e atributos, definir $p(a_m | y) = \mathbb{I}[a_m = a_m^y]$. A combinação de ambos os passos resulta:

$$p(a_m | x) = \sum_{k=1}^K p(a_m | y_k) p(y_k | x) \quad (3)$$

onde $p(a_m | y_k)$ é o atributo de classe predefinido e $p(y_k | x)$ é a classe de treino subsequente do classificador multi-classe.

Então para inferir as probabilidades subseqüentes do atributo $p(a_m | x)$ requer apenas uma multiplicação matriz-vetor. Depois disso, continuamos da mesma maneira que o DAP, classificando os exemplos de teste usando a Equação (2).

Animals with Attributes Dataset

- No seguimento dos trabalhos de Osherson, Wilkie e Kempt et al., os autores deste artigo recolheram imagens pertencentes a cada classe para formar o dataset AWA, que anteriormente apenas era formado apenas por conteúdo textual, relacionando as classes de animais com um conjunto de atributos.
- Os animais são unicamente caracterizados pelo seu vetor de atributos, o que possibilita a que este dataset sirva de base para a tarefa de incorporar o conhecimento humano num sistema de detecção de objetos.
- À data, o dataset conta com 37322 imagens de 50 classes diferentes, cada uma delas representada por 85 atributos.

Experiências

Setup

- **10 classes de teste:**

chimpanzee, giant panda, hippopotamus, humpback whale, leopard, pig, racoon, rat, seal.

- **Train/Test Split:**

Train: 80%

Test: 20%

- As experiências demonstraram que usando uma camada de atributos é possível construir um sistema de aprendizagem de detecção de objetos que não requer imagens de treino para as classes de teste.
- Para o **DAP** foi treinado um SVM não-linear para cada atributos binário a_1, \dots, a_M . Todas as SVMs são baseadas no mesmo kernel. O parâmetro C foi definido a 10. De maneira a obter as estimativas de probabilidades, o treino com as SVM foi feito usando apenas 90% dos exemplos de treino, usando o remanescente para estimar os parâmetros da curva sigmoid para Platt scaling, para converter o output das SVM nas estimativas de probabilidades. Na fase de teste foram aplicadas as SVMs treinadas com o Platt scaling a cada imagem de teste e feitas as previsões das classes de teste usando a Equação (2).
- Para o **IAP** foi usada regressão logística multi-class com o fator de regularização L2. Num outro paper, usaram SVMs one-versus-rest para cada classe de treino, usando 90/10 para as funções de decisão e para os coeficientes sigmoid para PLatt scaling. Na fase de teste, foi previsto um vetor das probabilidades de classes para cada imagem de teste. O vetor foi normalizado (L1) para que pudesse ser interpretado como a distribuição sobre as classes de treino. Isto dá diretamente um estimativa das classes $p(y_k|x)$ que são transformadas em atributos pela equação (3).

Resultados

- Tendo treinado os *predictors* para $p(a_m|x)$ na parte de treino do AWA, os vetores de atributos das classes de teste e a Equação (2) foram usados para levar a cabo a classificação multi-classe na parte de teste do dataset;
- O **DAP** conseguiu atingir melhor desempenho (40.5% *accuracy*);
- Tendo uma grande quantidade de informação de treino incluída, a performance pode ser significativamente maior.

Conclusão

- Foram propostos dois métodos para classificação baseada em atributos que resolve o problema de aprender com classes de treino e teste disjuntas transferindo informação entre as classes;
- Os atributos semânticos *per-class* são uma forma fácil de incluir o conhecimento humano num sistema de detecção de objetos;
- Foi também introduzido o *dataset Animals with Attributes*;

Explicação dos métodos, por outras palavras...

De uma maneira geral, os métodos apresentados neste artigo, o DAP e o IAP, têm duas fases importantes: numa primeira fase são previstos os atributos de uma imagem de entrada, e na segunda fase a classe é inferida procurando a classe que melhor se adequa ao conjunto de atributos.

O DAP aprende inicialmente classificadores probabilísticos para cada atributo e faz a previsão da classe combinando os resultados dos classificadores de atributos aprendidos, usando a Equação (2).

O IAP estima indiretamente as probabilidades dos atributos de uma imagem prevendo em primeiro lugar as probabilidades de cada classe de treino e depois multiplicando a matriz dos atributos de classe. As probabilidades dos atributos são obtidos através da Equação (3). A Equação (2) é usada depois para prever a classe desconhecida.

1. Métodos baseados em estimadores (Bayes). ↩
2. Em *machine learning* um classificador probabilístico é um classificador que é capaz de prever, para um dado exemplo de entrada, uma distribuição de probabilidade sobre o conjunto das classes, em vez de devolver apenas a classe que melhor classifica o exemplo de entrada. ↩