Predicting User Annoyance Using Visual Attributes

Gordon Christie¹  Amar Parkash²  Ujwal Krothapalli¹  Devi Parikh¹

¹Virginia Tech  ²Goibibo

Problem
• Computer vision algorithms make mistakes
• In human-centric applications, some mistakes are more annoying to users than others
• Where does this cost matrix come from?

Modeling User Annoyance
Notation
• Mistake: if an image from category $i$ is classified to be category $j$ (classification), or if an image from category $i$ is considered to be similar to an image from category $j$ (retrieval)
• We predict the cost (annoyance) of a mistake ($c_{ij}$) between an image from category $i$ and another from category $j$
• $2M$-dimensional feature vector ($d_{ij}$) for each category pair
• Attribute memberships: $M$-dimensional feature vector $a^m_i = \{0, 1\}$ to indicate that attribute $m$ is present in category $i$

Features
• Concatenate differences and similarities of attribute scores
  $d_{ij} = [d_{ij}^m \ d_{ij}^{\neg m}]$
  $d_{ij}^m = a^m_i \oplus a^m_j$
  $d_{ij}^{\neg m} = a^m_i \land a^m_j$

Experimental Setup
Data Collection
• Collect ground truth annoyance on Amazon Mechanical Turk (Range 1 - "Not annoyed at all" to 5 - "Very annoyed")

Evaluation – Annoyance Prediction
• Evaluate the regressor and ranker, which involves predicting a cost matrix for the current test categories
• Mean Squared Error (MSE) for the regressor
• Spearman Rank correlation coefficient for the ranker

Evaluation – Image Search
• Find the true mean and min annoyance of top $R$ retrieved results from the ranker over multiple runs

Results
Datasets
• Pubfig (faces)
  • 8523 images from 60 categories
  • Use a vocabulary of 63 out of 73 attributes
• SUN (scenes)
  • 1600 images from 80 categories used
  • Use 62 out of 102 attributes from Patterson and Hays

Annoyance Prediction
• We are able to predict annoyance more accurately than baselines (our approach = [ ], baselines = [ ])

Image Search
• Our approach makes less annoying mistakes in an image retrieval application

Conclusions
• Modeling differences and similarities in attribute-based representations allows us to predict annoyance of previously unseen mistakes more effectively than several baselines.
• Our approach allows for an improved user experience for image search.