Human Computation for Attribute and Attribute Value Acquisition

Edith Law, Burr Settles, Aaron Snook, Harshit Surana, Luis von Ahn, Tom Mitchell
School of Computer Science
Carnegie Melon University
edith@cmu.edu

Abstract

We introduce a human computation approach for acquiring object attributes and attribute values from the crowd, including a new game called Polarity.

1. Introduction

An important challenge in information retrieval is bridging the "semantic gap", which refers to the disconnect between the way that humans and machines represent and describe objects. The semantic gap prevents humans from expressing their information needs using natural language, and makes it difficult for machines to explain the relevance of the retrieved items to humans. Attributes help bridge this semantic gap. They are compoundable, making them extremely useful for constructing complex queries (e.g., "asian women with short hair, big eyes and high cheekbones") and identification (e.g., find an actor whose name you forgot, or an image that you have misplaced in a large collection). In recommendation systems, indexing objects by attributes makes it possible to explain why a particular item is chosen for the user (e.g., this song is recommended because it is "calm" and "sentimental", just like the other ones that you liked).

Attributes are also useful for learning. There has been a recent movement towards using an intermediate layer of human-understandable attributes for classification [12, 8, 4, 3, 7]. The idea is to build a two-layer classifier that maps image features first to a semantic code, then maps the semantic code to a set of classes. With the exception of [13], most works use a set of attributes that are fixed and manually curated. For example, Kumar et al. [7] manually created 65 attributes for face recognition, and paid workers on Mechanical Turk to obtain the attribute values for each image. The *Animal with Attributes* dataset was created using the 50 attributes proposed in [6, 11]. The outdoor scene datasets provided by [4, 3] uses a fixed set of 64 attributes, describing the objects' shape (e.g., "cylindrical"), parts (e.g., "has window") and material (e.g., "is shiny"). A few works use other sources of information, e.g., text corpus, to automatically characterize the visual attributes of objects [14, 2] without any human supervision.

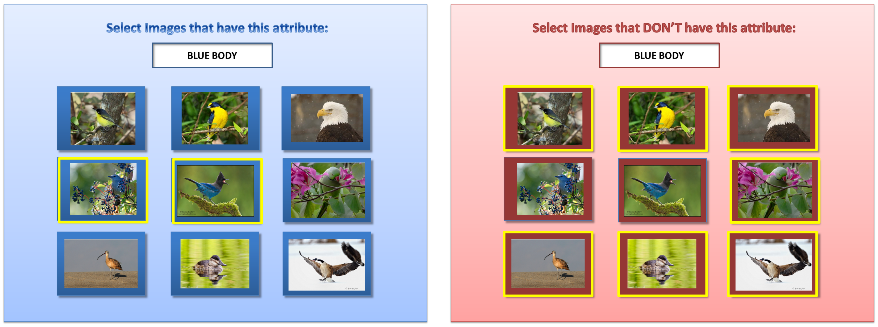
Human computation enables large-scale collection of semantic attributes of objects, which can then to use to index objects in a human-understandable way. The ESP Game mechanism, for example, has been used to collect millions of image tags, that are then used to power image search on the Web. Game mechanisms, however, is not one-size-fits-all. It has been noted that the ESP Game produces image tags that tend to be common and uninformative [16, 5]. This is a direct consequence of the output-agreement mechanism – needing to agree with his partner, a player's best strategy is to enter common tags that are likely to be entered by any person. Ad-hoc fixes, such as taboo words [15] or arbitrary restrictions on players' outputs (e.g., that the tag must start with the letter *a* [16]), do not seem to eliminate the problem completely. Because of the limitations of the output-agreement mechanism, it is necessary to invent new game mechanism in order to collect detailed, descriptive semantic attributes that can be used to make fine-grained distinctions between different categories and objects. In this work, we introduce a new game mechanism called *complementary-agreement mechanism* for collecting attributes and attribute values for images, and a game that implements this mechanism called Polarity.

2. Learning Attributes from the Crowd

Polarity is part of a larger integrated, machine-in-the-loop system for attribute learning [9]. Similar to Parikh and Krauman [13], Polarity will be integrated with a *active attribute acquisition* algorithm, which will intelligently choose a set of images for players to process during each round of the game. In this section, we will describe Polarity, its underlying game mechanism, and some of its properties.

Polarity: A Complementary-Agreement Mechanism

Consider a game for collecting object attributes, where two players are presented with a set of objects (e.g., images)



(a) "Positive Player"   (b) "Negative Player"

Figure 1. Polarity

and an attribute (e.g., "has red beak") and asked to click on the objects with that attribute. The output-agreement mechanism, in this case, would work poorly – if players are rewarded for agreement, then there is a simple cheating strategy where players click on everything and receive the maximum reward.

To solve this problem, we introduce a new game mechanism called *complementary-agreement* mechanism, where one of the players is asked to generate outputs that the other player is forbidden to enter. Polarity (Figure 1) is game that implements this mechanism. In the visual version of this game, two players are presented with a set of images. Players alternate between two roles – the "positive" player (Figure 1(a)) is asked to name an attribute and select images that the attribute describes, while the "negative player" (Figure 1(b)) is asked to select images that the named attribute does not describe. Players receive a joint score of $(|\mathcal{S}_p| \times |\mathcal{S}_n|) - c \cdot |\mathcal{S}_p \cap \mathcal{S}_n|$, where $\mathcal{S}_p$ is the set of entities selected by the positive player, $\mathcal{S}_n$ is the number of entities selected by the negative player, and c is the penalty for selections that overlap between the two players.

The complementary-agreement mechanism has some interesting properties. First, in a single round of the game, we are able to gather both the positive and negative examples of a given attribute. This allows rapid creation of datasets for training attribute classifiers. Second, since the entire set of the objects are revealed to the players (as opposed to games with hidden information, such as TagATune [10], where each player is given a partial set of the objects), we can gather attributes that *explicitly* distinguish between objects that are confusable. Finally, the game allows machine learners to propose new attributes and attribute values to be evaluated by the human players. This is a useful property for building a continuous attribute learning system that can monitor its own progress using human feedback.

2.1. Conclusion

Our short term goal is to collect image attributes using Polarity, and compare the performance of our two-layer classifier to existing works [8, 7, 4, 17, 1] which employ attributes collected via other means. Accompanying the poster will also be a demo of the visual version of the Polarity game, using images from five datasets, including Animals with Attributes [8], PubFig [7], aPascal [4], CUB-200 [17] and LeafSnap [1].

References

[1] Leafsnap website. http://leafsnap.com. 2

[2] T. Berg, A. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *ECCV*, 2010. 1

[3] A. Farhadi, I. Endres, and D. Hoiem. Attribute-centric recognition for cross-category generalization. In *CVPR*, 2010. 1

[4] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *CVPR*, 2009. 1, 2

[5] S. Jain and D. Parkes. A game-theoretic analysis of games with a purpose. In *WINE '08: Proceedings of the 4th International Workshop on Internet and Network Economics*, pages 342–350, Berlin, Heidelberg, 2008. Springer-Verlag. 1

[6] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, 2006. 1

[7] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2009. 1, 2

[8] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009. 1, 2

[9] E. Law. *Attribute Learning using Joint Human and Machine Computation: A Thesis Proposal*. PhD thesis, Carnegie Mellon University, April 2011. 1

[10] E. Law and L. von Ahn. Input-agreement: A new mechanism for collecting data using a human computation game. In *submission*. 2

[11] D. N. Osherson, J. Stern, O. Wilkie, M. Stob, and E. E. Smith. Default probability. 15(2), 1991. 1

[12] M. Palatucci, D. Pomerleau, G. Hinton, and T. Mitchell. Zero-shot learning with semantic output codes. In *Neural Information Processing Systems (NIPS)*, December 2009. 1

[13] D. Parikh and K. Krauman. Interactively building a discriminative vocabulary of nameable attributes. In *CVPR*, 2011. 1

[14] M. Rohrbach, M. Stark, G. Szarvas, I. Gurevych, and B. Schiele. What helps where – and why? semantic relatedness for knowledge transfer. In *CVPR*, 2010. 1

[15] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI*, pages 319–326, 2004. 1

[16] I. Weber, S. Robertson, and M. Vojnovic. Rethinking the esp game. In *CHI*, pages 3937–3942, 2009. 1

[17] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010. 2