Understanding User's Intention in Semantic Based Image Retrieval: Combining Positive and Negative Examples

Meriem Korichi[1]([✉]), Mohamed Lamine Kherfi[2], Mohamed Batouche[1], Zineb Kaoudja[2], and Hadjer Bencheikh[2]

[1] Computer Science Department, University of Constantine 2 - Abdelhamid Mehri, 25 000 Constantine, Algeria
meriemkorichi@gmail.com, mcbatouche@gmail.com
[2] Department of Computer Science and Information Technology, University of Ouargla - Kasdi Merbah, 30000 Ouargla, Algeria
Mohammedlamine.Kherfi@uqtr.ca, zinebkaoudja@gmail.com, bencheikh1991@gmail.com

Abstract. Understanding user's intention is at the core of an effective images retrieval systems. It still a significant challenge for current systems, especially in situations where user's needs are ambiguous. It is in this perspective that fits our study.

In this paper, we address the challenge of grasping user's intention in semantic based images retrieval. We propose an algorithm that performs a thorough analysis of the semantic concepts presented in user's query. The proposed algorithm is based on an ontology and takes into account the combination of positive and negative examples. The positive examples are used to perform generalization and the negative examples are used to perform specialization which considerably decrease the two famous problems of image retrieval: noise and miss.

Our algorithm processed in two steps: in the first step, we deal only with the positive examples where we will generalize the query from the explicit concepts to infer the others hidden concepts desired by the user. whereas the second step deal with the negative examples to refine results obtained in the first step. We created an image retrieval system based on the proposed algorithm. Experimental results show that our algorithm could well capture user's intention and improve significantly precision and recall.

Keywords: Image retrieval · Grasping user's intention
Positive examples · Negative examples · Ontology

1 Introduction

Image retrieval is a growing field. Recently, it has witnessed an explosion of personal and professional image collections especially with the development of new technologies that allow users share their images via internet. There was an urgent need for automatic tools that organizing these large quantities of images. Such tools can help users to find the desired images within a reasonable time. These tools are called image retrieval engines.

Image retrieval systems can be classified into two main categories: the first category called Content-Based Image Retrieval (CBIR). In this category, the user is generally asked to formulate a visual query (image for example) that correspond to what he is looking for. The research is performed by measuring similarities between the low levels features of the query and the images of the collection. Various visual features, including both global features [1–3] (e.g., color, texture, and shape) and local features (e.g., SIFT keypoints [4–6]) have been studied for CBIR.

The second category which exploits the semantic concepts associated with images is called Semantic-Based Image Retrieval: SBIR. The user formulates his/her query with textual terms that express his/her needs. In this case, the search is performed by comparing between these terms and textual annotations that represent the images in the collection by the use of semantic tools. Among the tools that are often used to represent the semantics, there are "ontologies" [7].

There is another secondary category where the user formulates his/her query with visual example like CBIR but the research is carried out on the basis of the semantics associated with those images instead of the low levels features [8]. This category try to overcome the drawbacks of the two previous categories by avoiding user to identify keywords and letting the images speak for themselves.

Whatever the category of the image retrieval system, Their main objective is to understand user's needs and answer him/her accurately. However, these needs may vary from one person to another, and even from a given user at different times; therefore, the success of an image retrieval system depends on how accurate the system understands the intention of the user when he/she formulates a query [9].

Creating the query is a difficult problem. It arises two important challenges for both the user and the retrieval system [9]: Firstly, for the user the challenge is how to express his/her needs accurately. Secondly, for the system the challenge is how to understand what the user wants based on the query that he/she formulated.

The difficulty in dealing with these two challenges lead to many problems in image retrieval, such as noise and miss. Noise can be defined as the set of retrieved images which do not correspond to what the user wants [10] whereas, miss is the set of images corresponding to what the user wants which have not been retrieved.

In this paper, our objective is to improve the efficiency of image retrieval by decrease noise and miss via understanding user's intention. We will be interested by the systems of the last category where the user formulates a visual query using

images and the retrieval system exploits the semantic concepts associated with these images.

We propose an algorithm that performs a thorough analysis of the semantic concepts presented in the user's query. Our algorithm is processed in two steps: in the first step, we will perform a generalization from the explicit concepts to infer the other hidden concepts that the user can't express them which decrease the miss problem. In the second step, we ask user to select negative examples to refine results obtained in the first step. In this case the images similar to those of the negative examples will be rejected, which reduces the noise. At the same time, the rejected images are replaced by others correspond to what the user wants, so miss will be decrease also.

The rest of the paper is organized as follows: Sect. 2 briefly reviews previous related work. Section 3 present the principle of our algorithm, then, we will present our data and our working hypotheses in Sect. 4. Section 5 details our algorithm. An experimental evaluation of the performance of our algorithm is presented in Sect. 6. Finally, Sect. 7 presents a conclusion and suggests some possibilities for future research.

2   Related Works

User is at the centre of any image retrieval system. One of the primary challenges for these systems is to grasp user intent given the limited amount of data from the input query [11]. In order to solve this challenge, several methods has been proposed. One way is keyword expansion where the user's original query is augmented by new concepts with a similar meaning to decrease the miss problem.

Keyword expansion was implemented by different techniques. Some of them [12–14] are typically based on dictionaries, thesauri, or other similar knowledge models such as ontology. It is noted by many researchers that the use of an ontology for query expansion is advantageous especially if the query words are disambiguated [11,15]. Note that query expansion is very effective, provided that initial query perfectly matches the user's need [16], However, this is not always the case, especially in situation where the user has a difficulty to express them. In this case, the extensions of the query can lead to more inappropriate results than the initial query because of irrelevant concepts which leads to significantly increase noise [17].

Another way to grasp user's intention is relevance feedback [10,18–20] and pseudo relevance feedback. In Relevance feedback technique, the user choose from the returned images those that he/she finds relevants (positive examples) and those that he/she finds not relevants (negative examples) then a query-specific similarity metric was learned from the selected examples.

In pseudo relevance feedback [21,22], the query is expanding by taking the top N images visually most similar to the query image as positive examples.

For many researchers [23–25], The task of learning concepts from user's query to understand his/her intent has seen as a supervised learning problem in which the classifiers search through a hypothesis space to find an adequate hypothesis that will make good predictions. Several algorithms of supervised learning were proposed [26–29] such us: Multi-Class Support Vector Machines (SVM), Naive Bayes Classifier, Maximum Entropy (MaxEnt) classifier, Multi-Layer Perceptron (MLP).

Recent search [25], proposed an ensembling multiple classifiers for detecting user's intention instead of focusing on single classifiers.

This problem has an analogue in the cognitive sciences which called human generalization behavior and inductive inference.
Bayesian models of generalization [30–33] are models from cognitive science that focus on understanding the phenomena of how to generalize from few positive examples. They has been remarkably successful at explaining human generalization behavior in a wide range of domains [33], however their success is largely depend on the reliability of the initial query.
The main contribution of our work is to propose an algorithm to perform the challenge of grasping user's intention and consequently decrease miss and noise. Our algorithm refines the initial query formulated by the user to infer the relevant concepts. It uses an ontology and exploits its semantic richness which allowed us to extract the hidden concepts that the user could not express them explicitly. Then we can improve the obtained results through the relevance feedback with negative examples.

3   Principle of Our Algorithm

In medicine, doctors apply a characteristic rule to summarize the symptoms of a specific disease, whereas, to distinguish one disease from others, they apply a discrimination rule to summarize the symptoms that discriminate this disease from others. This principle is exactly what we will use in our algorithm to understand user's intention from a given query. We perform generalization using positive examples and specialization using negative examples. This principle is applied in content-based image retrieval [10] as follows:

– For positive examples, a concentrated characteristic is important and a scattered characteristic is not important.
– For positive and negative examples, a discriminated characteristic is important and a non-discriminated characteristic is not important.

  We carry out a similar work with semantic concepts as follows:

1. For positive examples: with a query that contains several images:
   (a) A concept which is salient in the majority of query's images is important.

(b) A concept that appears in some images and does not appear in the others, in this case, we cannot judge immediately that it is not important. However, it can be classified into three types:

    i. Either, it has concepts that reinforces it in the other images via direct or indirect relationships (for example: apple - fruit - banana), and so it is important, or at least the common concept (here it is the ancestor Fruit) is important.

    ii. Either it has concepts that contradicts it in the other images, directly or indirectly, and therefore it is not important.

    iii. Either there is no concept that reinforces it or a concept that contradicts it in the other images, in this case, it can either be considered with moderate importance, or neglected.

2. For positive and negative examples:

    (a) The common concepts between the two sets (positive and negative) are to be neglected. common concept mean either the same concept or concepts that have a positive relations between them as it is illustrated in the case of positive example only.

    (b) Discriminant concepts (directly or indirectly through relationships) between the two sets are important. If the concept is present in the positive set, it must be sought; and if it is present in the negative set, it must be dismissed.

4 Data and Hypothesis

4.1 Query Formulation

We ask user to select from displayed images the maximum number of images that correspond to what he/she is looking for. It is the set of concepts that annotate the selected images that will constitute our initial query.

4.2 Ontology Creation

Our algorithm use an ontology to exploit its semantic richness and apply the principle describe above. We have chosen the domain of "nature" for our ontology. The first step of the ontology creation is to choose a collection of images representative of "Nature" domain. The ontology will then be used to annotate the images in this collection. Relationships in our ontology are classified into three categories:

1. Positive relationships (reinforcement relationship): means that the presence of such concept implies the presence of such other concept, for example: presence of concept "Dune" implies the presence of concept "Desert".
2. Negative relationships (contradiction relationships): means that the presence of such concept generally implies the absence of such other concept, for example: presence of the concept "Snow" implies the absence of concept "Desert".

3. Neutral relationships: means that the relations between concepts are null (ie neither positive nor negative).

The other steps of the ontology creation are the following:

1. Identify the different concepts presents in the images collection: for example: desert, camel, dune, snow, etc.
2. Calculate the implications weights between concepts that illustrate the degree of semantic correlation between these concepts.
   For each concept, the probability of having the other concepts was calculated as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

   Where $P(A \mid B)$ means the probability of having a concept A knowing that we are in the presence of the concept B.
   Because our challenge is not the ontology creation, we have calculated this probability manually. Firstly, we started the search with the concept B using the retrieval engine "Google". The obtained images represent P(B). In this collection, we calculate the number of images that represents the concept A which gives us $P(A \cap B)$.
3. Represent the ontology by a value oriented graph where concepts represent the nodes and the arcs are the semantic relationships. Our graph can be seen as a probabilistic graph.
4. Represent the relationships between concepts with a relevance matrix. This latter is a square matrix whose lines and columns are concepts and the value in box M [i, j] represents the probability of the relation between the concept i and the concept j.

A weight between two concepts models a type of semantic relationship between them.

* The weights with the value 0: mean that the two concepts can never be found in the same image so this weight models a negative relationship;
* [0,01 0, 49]: meant that there are times when the two concepts can be found in the same image and sometimes it cannot be found. Which models a neutral relationship;
* [0.5 1]: means that the two concepts are often found in the same image so this weight models a positive relationship.

5 Our Algorithm

Our algorithm is processed in two steps. We called the first step: "Generalization algorithm" (see Algorithm 1) and the second step "Specialization algorithm" (see Algorithm 2).

Algorithm 1. Generalization

Background: a set of an annotated images D
Preprocess: Extract the list of semantic concepts (LCpt) that annotate the positive examples
Input: list of concept LCpt =C1, C2, ... Cn, relevance matrix M, ontology G
Output: L1 list of relevant concepts (explicit and hidden)
  for all C ∈ LCpt **do**
    compute weight (C): the number of occurrences of C in Q
    compute P(C) whith
    W ← weight (C)
    NB ← Number of images in Q

$$P(C) = \frac{W}{NB}$$

    if P(C) ≥ 0.5 **then**
      add (C) to L1 (the list of accepted concepts)
    else
      add (C) to L2 (the list of concepts to be verified)
    end if
  end for
  for all C1 ∈ L2 **do**
    for all C2 ∈ LCpt **and** C1 ≠ C2 **do**
    Sum = Sum + M[C1,C2]

$$Avg = \frac{Sum}{length(LCpt) - 1}$$

    if Avg ≥ 0.5 **then**
      add (C1) to L1
    end if
    end for
  end for
  for all C1 ∈ LCpt **do**
    for all C2 ∈ G **do**
      if M[C1,C2] ≥ 0.5 **then**
        add (C2) to Hid (the list of hidden concepts)
      end if
    end for
  end for
  for all C ∈ Hid **do**
    compute wght (C):the number of occurrences of C in Hid
    compute Moy(C) whith
    Wh ← wght (C)
    Nb ← length (LCpt)

$$Moy(C) = \frac{Wh}{Nb}$$

    if Moy(C) ≥ 0.5 **then**
      add (C) to L1
    end if
  end for

Algorithm 2. Specialization

    Background: a set of an annotated images D (search results using L1 as query)
    Preprocess: Extract the list of semantic concepts (LNE) that annotate the negative
    examples
Input: list of relevant concepts L1, list of negative concepts LNE, relevance matrix
    M, ontology G
Output: L3 list of undesired concepts.
    Apply the Generalization algorithm to the list LNE, as results we obtain a an initial
    list of undesired concepts (L3).
    for all $C \in$ L1 **do**
      if $C \in$ L3 **then**
        delete C from L3
      end if
    end for
    for all $Ci \in$ L3 **do**
      for all $Cj \in$ L1 **do**
        Som = Som + M[Ci,Cj]

$$Moyn = \frac{Som}{length(L1)}$$

      if Moyn $\geq$ 0.5 **then**
        delete Ci from L3
      end if
    end for
    end for
    Display results by deleting all images where the concepts of l3 are appeared

---

6 Experimental Evaluation

6.1 Images Collection

To be able to apply our algorithm, we need an ontology and an images collection annotated with the concepts of this ontology. We have chosen the domain of "nature" for our ontology. Then, we have used Google images as an image retrieval engine to collect images for each concept in the ontology. The top 200 returned images of each query are crawled and manually labeled to construct the experimental data set.

6.2 Experiments

We implemented an image retrieval system that integrate our algorithm. We invited thirty users to participate in our experiments. We displayed them a set of images randomly selected from the images collection. Users can ask to see more images if the presented images do not reflect to what they wants.

    To evaluate the performance of our algorithm, we carried out a comparison between: our algorithm, the Bayesian model of generalization used in [33] and the

standard matching method (without generalization). As performance criteria, we use the precision and the recall measures.

First Experiment. The aim of this experiment is to evaluate the effectiveness of the first step of our algorithm (Generalization algorithm) to increase recall compared to the Bayesian model of generalization used in [33] and the standard matching method (without generalization). We asked users to make different queries, and then we calculated the recall of the three models. Fig. 1 illustrate the obtained results from 16 queries.
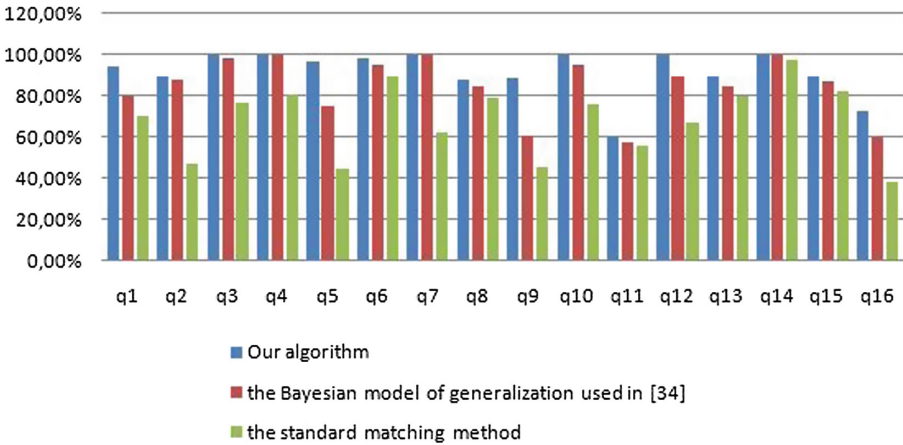

Fig. 1. The recall of the three models

We can see that the recall is considerably increased with our approach and then consequently the miss problem is decreased. Compared to our algorithm, the success of the Bayesian model of generalization is largely depend on the reliability of the examples present in the query to good reflect the user's intention which is not always the case because sometimes the images of the query may composed of many objects, however the user is interest only with some of them. Our algorithm overcomes this problem by refining the initial query of the user.

Second Experiment. The aim of this experiment is to illustrate the effect of the negative examples to decease noise by improving the accuracy. We asked users to formulate queries using the Generalization algorithm and we calculated the accuracy then we asked him to select negative images and restart the search. the obtained results are illustrated in Fig. 2.

It is clear that the use of negative examples increase the accuracy. Indeed, after obtaining the results of a given query, the user can keep the positive images and enrich the query by including some undesired images as a negative examples.
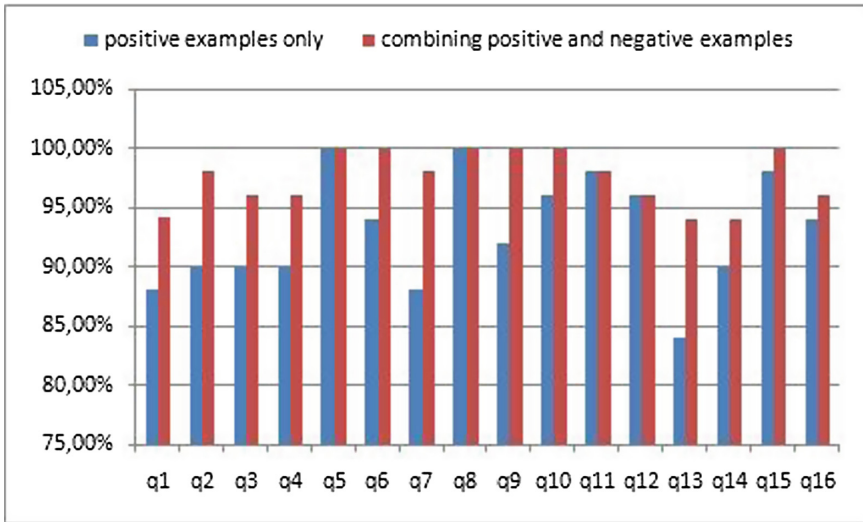
Fig. 2. The effect of negative examples to increase accuracy.

This implies that images similar to those of the negative example will be rejected, which reduces the noise. At the same time, the rejected images are replaced by others corresponding to what the user wants, so silence will be diminished.

7   Conclusion

In this paper, we proposed an algorithm that process in two steps: we called the first step "Generalization algorithm" and the second step "specialization algorithm". The objective of this algorithm is to improve the semantic image retrieval quality by reducing the famous problems of image retrieval: noise and miss via understanding user's intention. Our algorithm is based on the use of an ontology which allows us to exploits it's semantic richness. We implemented an image retrieval system that integrate the proposed algorithm. Experimental evaluation verify the efficiency of our approach in comparison to the others proposed in the state-of-the-art. We have shown that the combination of positive and negative examples improve significantly precision and recall. The perspective of our work is to automate the ontology creation.

References

1. Lowe, D.G.: Object recognition from local scale-invariant features. In: IEEE ICCV (1999)
2. Torralba, A., Murphy, K., Freeman, W., Rubin, M.: Context-based vision system for place and object recognition. In: Proceedings of the International Conference on Computer Vision (2003)

3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (2005)

4. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)

5. Celik, C., Bilge, H.S.: Content based image retrieval with sparse representations and local feature descriptors: a comparative study. Pattern Recogn. **68**, 1–13 (2017)

6. Guillaumin, M., Mensink, T., Verbeek, J., Schmid, C.: TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation. In: IEEE ICCV, pp. 309–316 (2009)

7. Liaqat, M., Khan, S., Majid, M.: Image retrieval based on fuzzy ontology. Multimed. Tools Appl. **76**, 22623–22645 (2017)

8. Rasiwasia, N., Moreno, P.J., Vasconcelos, N.: Bridging the gap: query by semantic example. IEEE Trans. Multimed. **9**(5), 923–938 (2007)

9. Kherfi, M.L.: Review of human-computer interaction issues in image retrieval. In: Pinder, S. (ed.) Advances in Human Computer Interaction. InTech (2008). https://doi.org/10.5772/5929

10. Kherfi, M.L., Ziou, D., Bernardi, A.: Combining positive and negative examples in relevance feedback for content-based image retrieval. J. Vis. Commun. Image Represent. **14**(4), 428–457 (2003)

11. Tang, X., Liu, K., Cui, J., Wen, F., Wang, X.: Intentsearch: capturing user intention for one-click internet image search. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1342–1353 (2012)

12. Bilotti, M., Katz, B., Lin, J.: What works better for question answering: stemming or morphological query expansion? In: Proceedings of the Information Retrieval for Question Answering (IR4QA) Workshop at SIGIR 2004 (2004)

13. Navigli, R.: Word sense disambiguation: a survey. ACM Comput. Surv. **41**(2), 1–69 (2009)

14. Azizan, A., Bakar, Z.A., Noah, S.A.: Analysis of retrieval result on ontology-based query reformulation. In: Proceedings of the 1st International Conference on Computer, Communication and Control Technology, I4CT 2014, pp. 244–248 (2014)

15. Zha, Z.-J., Yang, L., Mei, T., Wang, M., Wang, Z.: Visual query suggestion. In: Proceedings of the 17th ACM International Conference on Multimedia, MM 2009, vol. 6, no. 3, p. 15 (2009)

16. Fergus, R., Perona, P., Zisserman, A.: A visual category filter for google images. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 242–256. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24670-1_19

17. Park, G., Baek, Y., Lee, H.-K.: Majority based ranking approach in web image retrieval. In: Bakker, E.M., Lew, M.S., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 111–120. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45113-7_12

18. Deng, J., Berg, A.C., Fei-Fei, L.: Hierarchical semantic indexing for large scale image retrieval. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (2011)

19. Glowacka, D., Shawe-Taylor, J.: Image Retrieval with a Bayesian Model of Relevance Feedback (2016)

20. Tao, D., Tang, X., Li, X., Wu, X.: Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1088–1099 (2006)

21. Torjmen, M., Pinel-Sauvagnat, K., Boughanem, M.: Using pseudo-relevance feedback to improve image retrieval results. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 665–673. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85760-0_85
22. Yan, R., Hauptmann, A., Jin, R.: Multimedia search with pseudo-relevance feedback. In: Bakker, E.M., Lew, M.S., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 238–247. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45113-7_24
23. Kang, I.-H., Kim, G.: Query type classification for web document retrieval. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, SIGIR 2003, pp. 64–71. ACM, New York (2003)
24. Lee, U., Liu, Z., Cho, J.: Automatic identification of user goals in web search. In: Proceedings of the 14th International Conference on World Wide Web, WWW 2005, pp. 391–400. ACM, New York (2005)
25. Figueroa, A., Atkinson, J.: Ensembling classifiers for detecting user intentions behind web queries. IEEE Internet Comput. **20**(2), 8–16 (2016)
26. Krammer, K., Singer, Y.: On the algorithmic implementation of multi-class svms. Proc. J. Mach. Learn. Res. **2**, 265–292 (2001)
27. Hsu, C.-W., Lin, C.-J.: A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Networks **13**(2), 415–425 (2002)
28. Nigam, K., Lafferty, J., Mccallum, A.: Using maximum entropy for text classification
29. Foody, G.M.: Hard and soft classifications by a neural network with a non-exhaustively defined set of classes. Int. J. Remote Sens. **23**(18), 3853–3864 (2002)
30. Tenenbaum, J.B., Griffiths, T.L.: Generalization, similarity, and Bayesian inference. Behav. Brain Sci. **24**, 629–630 (2001)
31. Tenenbaum, J.B., Xu, F.: Word learning as Bayesian inference. In: Proceedings of the 22nd Annual Conference of the Cognitive Science Society (2000)
32. Abbott, J.: Constructing a hypothesis space from the Web for large-scale Bayesian word learning. Proc. Annu. Meet. Cogn. Sci. Soc. **34**, 54–59 (2012)
33. Jia, Y., Abbott, J., Austerweil, J.L., Griffiths, T.L., Darrell, T.: Visual concept learning: combining machine vision and Bayesian generalization on concept hierarchies. In: Advances in Neural Information Processing Systems 27 (NIPS 2013), vol. 1, no. 1, pp. 1–9 (2013)