NRC Publications Archive Archives des publications du CNRC

An extension of the aspect PLSA model to active and semi-supervised learning for text classification

Krithara, Anastasia; Amini, Massih-Reza; Goutte, Cyril; Renders, Jean-Michel

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

https://doi.org/10.1007/978-3-642-12842-4_22

Artificial Intelligence, pp. 183-192, 2010-06-01

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=f2c2f8f4-bb07-4f30-bcba-064c39ff0268>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=f2c2f8f4-bb07-4f30-bcba-064c39ff0268>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

An Extension of the Aspect PLSA Model to Active and Semi-Supervised Learning for Text Classification

Anastasia Krithara¹, Massih-Reza Amini², Cyril Goutte²,
and Jean-Michel Renders³

¹ National Center for Scientific Research (NCSR) 'Demokritos', Athens, Greece

² National Research Council Canada, Gatineau, Canada

³ Xerox Research Centre Europe, Grenoble, France

Abstract. In this paper, we address the problem of learning aspect models with partially labeled examples. We propose a method which benefits from both semi-supervised and active learning frameworks. In particular, we combine a semi-supervised extension of the PLSA algorithm [11] with two active learning techniques. We perform experiments over four different datasets and show the effectiveness of the combination of the two frameworks.

1 Introduction

The explosion of available information during the last years has increased the interest of the Machine Learning (ML) community for different learning problems that have been raised in most of the information access applications. In this paper we are interested in the study of two of these problems which are the handling of partially labeled data and the modeling of the generation of textual observations.

Semi-Supervised Learning (SSL) has emerged in the Machine Learning community in the late 90's. Under this framework, the aim is to establish a decision rule based on both labeled and unlabeled training examples. To achieve this goal, the decision rule is learned by simultaneously optimizing a supervised empirical learner on the labeled set, while respecting the underline structure of the unlabeled training data in the input space.

In the same vein, Active Learning addresses also the issue of the annotation burden, but from a different perspective. Instead of using all the unlabeled data together with the labeled ones, it tries to minimize the annotation cost by labeling as few examples as possible and focussing on the most useful examples. Different types of active learning methods have been introduced in the literature, such as uncertainty-based methods ([13,21,3]), expected error minimization methods ([10,19,7]) and query by committee methods ([20,8,17,6]).

By combining semi-supervised and active learning, an attempt is made to benefit from both frameworks to address the annotation burden problem. The semi-supervised learning component improves the classification rule and the measure

of its confidence, while the active learning queries for labelling the most relevant and potentially useful examples.

On the other hand, new generative aspect models have recently been proposed which aim to take into account data with multiple facets. In this class of models, observations are generated by a mixture of aspects, or topics, each of which being a distribution over the basic features of the observations. Aspect models ([9,1]) have been successfully used for various textual information access and image analysis tasks such as document clustering and categorization or scene segmentation. In many of these tasks, acquiring the annotated data necessary to apply supervised learning techniques is a major challenge, especially in very large data sets. These annotations require humans who can understand the scene or the text, and are therefore very costly, especially in technical domains.

In this paper, we explore the possibility to learn such models with the help of the unlabeled examples, by combining SSL and active learning. This work is the continuation of, and builds on earlier work on SSL for PLSA [11]. In particular, the combination of the SSL variant of PLSA with two active learning techniques.

2 Combining SSL and Active Learning

The idea of combining active and semi-supervised learning was first introduced by [15]. The idea is to integrate an EM algorithm with unlabeled data into an active learning, and more particularly in a query by committee (QBC) method. The committee members are generated by sampling classifiers according to the distribution of classifier parameters specified by the training data. In [16], *Co-EMT* is proposed. This algorithm combines Co-Testing and Co-EM. As opposed to Co-Testing algorithm, which learns hypotheses h_1 and h_2 based only on the labeled examples, Co-EMT learns the two hypotheses by running Co-EM on both labeled and unlabeled examples. Then, in the active learning step, it annotates the example on which the predictions of h_1 and h_2 are the most divergent, that is the example for which h_1 and h_2 have an equally strong confidence at predicting a different label. [24] also presents a combination of semi-supervised and active learning using Gaussian fields and harmonic functions. [23] presented the so-called method *Semi-Supervised Active Image Retrieval (SSAIR)* for a different task of relevance feedback. The method was inspired by co-training [2] and co-testing [17], but instead of using two sufficient but redundant views of the dataset, it employs two different learners on the same data. In the context of multi-view active learning, [18] proposed a method which combines semi-supervised and active learning. The first step uses co-EM with naive Bayes as the semi-supervised algorithm. They present an approximation to co-EM with naive Bayes that can incorporate user feedback almost instantly and can use any sample-selection strategy for active learning.

Why the combination should work? The combination of both semi-supervised and active learning appears to be particularly beneficial in reducing the annotation burden for the following reasons:

1. It constitutes an efficient way of solving the exploitation/exploration problem: semi-supervised learning is more focused on exploitation, while active learning is more dedicated to exploration. Semi-supervised learning alone may lead to poor performance in the case of very scarce initial annotation. It strongly suffers from poorly represented classes, while being very sensitive to noise and potential instability. On the other hand, active learning alone may spend too much time querying useless examples, as it can not exploit the information given by the unlabeled data.
2. In the same vein, it may alleviate the data imbalance problem due to each method separately. Semi-supervised learning tends to over-weight easy-to-classify examples that will dominate the process, while active learning has the opposite strategy, resulting in exploring more deeply the hard-to-classify examples [22].
3. Semi-supervised learning is able to provide a more motivated estimation of the confidence score associated to the class prediction for each example, taking into account the whole data set, including the unlabelled data. As a consequence, active learning based on these better confidence scores is expected to be more efficient.

3 Semi-Supervised PLSA with a Mislabeling Error Model

In this section we present the semi-supervised variant of the Probabilistic Latent Semantic Analysis (PLSA) model which is used in combination with active learning. This method incorporate a misclassification error model (namely the *ssPLSA-mem*) [11]. We assume that the labeling errors made by the generative model for unlabeled data come from a stochastic process and that these errors are inherent to semi-supervised learning algorithms. The idea here is to characterize this stochastic process in order to reduce the labeling errors computed by the classifier for unlabeled data in the training set.

We assume that for each unlabeled example $x \in X_u$, there exists a perfect, true label y , and an imperfect label $\tilde{y}$, estimated by the classifier. Assuming also that the estimated label is dependent on the true one, we can model these labels by the following probabilities:

$$\forall(k, h) \in \mathcal{C} \times \mathcal{C}, \beta_{kh} = P(\tilde{y} = k | y = h) \quad (1)$$

subject to the constraint that $\forall h, \sum_k \beta_{kh} = 1$.

The underlying generation process associated to this latent variable model for unlabeled data is:

- Pick an example x with probability $P(x)$,
- Choose a latent variable α according to its conditional probability $P(\alpha | x)$
- Generate a feature w with probability $P(w | \alpha)$

- Generate the *latent* class y according to the probability $P(y | \alpha)$
- The imperfect class label $\tilde{y}$ is generated with probability $\beta_{\tilde{y}|y} = P(\tilde{y} | y)$

The values of $P(y|\alpha)$ depend on the value of latent topic variable α . The cardinal of α is given. The number of latent topics α per class is also known for both labeled and unlabeled examples. We initialize by forcing to zero the $P(y|\alpha)$ for the latent topic variables α which do not belong to the particular class y . These values remain fixed. In other words, we perform *hard clustering*. We have to note that the hard clustering is done for each class separately, since in each class (y) the corresponding feature examples may aggregate to several clusters.

In algorithm 1 the estimation of model parameters

$$\Phi = \{P(\alpha | x), P(w | \alpha), \beta_{\tilde{y}|y} : x \in \mathcal{X}, w \in \mathcal{W}, \alpha \in A, y \in \mathcal{C}, \tilde{y} \in \mathcal{C}\}$$

is described. This algorithm is an EM-like algorithm.

With $n(x, w)$ we denote the frequency of the feature w in the example x . For more information about this model, please refer to [11].

Algorithm 1. ssPLSA-mem

Input :

- A set of partially labeled data $\mathcal{X} = X_l \cup X_u$,
- Random initial model parameters $\Phi^{(0)}$.
- $j \leftarrow 0$
- Run a simple PLSA algorithm for the estimation of the initial $\tilde{y}$ for each example

repeat

E-step: Estimate the latent class posteriors

$$\pi_\alpha(w, x, y) = \frac{P(\alpha|x)P(w|\alpha)P(y|\alpha)}{\sum_\alpha P(\alpha|x)P(w|\alpha)P(y|\alpha)}, \text{ if } x \in \mathcal{X}_l$$

$$\tilde{\pi}_\alpha(w, x, \tilde{y}) = \frac{P(\alpha|x)P(w|\alpha) \sum_y P(y|\alpha)\beta_{\tilde{y}|y}}{\sum_\alpha P(\alpha|x)P(w|\alpha) \sum_y P(y|\alpha)\beta_{\tilde{y}|y}}, \text{ if } x \in \mathcal{X}_u$$

M-step: Estimate the new model parameters $\Phi^{(j+1)}$

 by maximizing the complete-data log-likelihood

$$P^{(j+1)}(w|\alpha) \propto \sum_{x \in X_l} n(w, x) \pi_\alpha^{(j)}(w, x, y(x)) + \sum_{x \in X_u} n(w, x) \tilde{\pi}_\alpha^{(j)}(w, x, \tilde{y}(x))$$

$$P^{(j+1)}(\alpha|x) \propto \sum_w n(w, x) \times \begin{cases} \pi_\alpha^{(j)}(w, x, y(x)), & \text{for } x \in X_l \\ \tilde{\pi}_\alpha^{(j)}(w, x, \tilde{y}(x)), & \text{for } x \in X_u \end{cases}$$

$$\beta_{\tilde{y}|y}^{(j+1)} \propto \sum_w \sum_{x \in X_u} n(w, x) \sum_{\alpha | \alpha \in y} \tilde{\pi}_\alpha^{(j)}(w, x, \tilde{y})$$

$j \leftarrow j + 1$

until convergence of the complete-data log-likelihood ;

Output : A generative classifier with parameters Φ

4 Active Learning

In this section, we extend the presented semi-supervised model, by combining it with two active learning methods. The motivation is to try to take advantage of the characteristics of both frameworks. In both models, we choose to annotate the less confident example. Their difference lies on the measure of confidence they use.

Margin Based Method. The first active learning method (the so-called margin based method) chooses to annotate the example which is closer to the classes' boundaries [12]. The latter gives us a notion of confidence the classifier has on the classification of these examples. In order to measure this confidence we use the following class-entropy measure for each unlabeled example:

$$B(x) = - \sum_y P(y|x) \log P(y|x), \text{ where } x \in \mathcal{X}_u \quad (2)$$

The bigger the B is, the less confident the classifier is about the labeling of the example. After having selected an example, we annotate it and we add it to the initial labeled set $\mathcal{X}_l$. More than one examples can be selected at each iteration. The reason is that, especially for classification problems with a big amount of examples and many classes, the annotation of only one example at a time, can be proved time-consuming, as a respectful amount of labeled examples will be needed in order to achieve a good performance. If we choose to do the latter, it is not wise to choose examples that are next to each other, as they cannot give us significantly more information than each of them does. As a result, it is better to choose, for instance, examples with big class-entropy which have been given different labels. That way the classifier can get information about different classes and not only for a single one.

Entropy Based Method. Based on the method presented in [5], we calculate the entropy of the annotation of the unlabeled data, during the iterations of the

Algorithm 2. Combining ssPLSA and Active Learning

Input : A set of partially labeled examples $\mathcal{X} = \mathcal{X}_l \cup \mathcal{X}_u$

repeat

- Run the ssPLSA algorithm (and calculate the $P(y|x)$)
- Estimate the confidence of the classifier on the unlabeled examples
- Choose the example(s) with low confidence
(if we choose more than one example to label, we choose examples with have been classified into different classes), annotate them and add them in the labeled dataset $\mathcal{X}_l$

until a certain number of queries or a certain performance ;

Output : A generative classifier

model. This method can be seen as a query by committee approach, where, in contrast to the method of [5], the committees here are the different iterations of the same model.

In contrast to the margin based method presented previously, the current one does not use the probabilities $P(y|x)$ of an example x to be assigned the label y , but instead, it uses the deterministic votes of the classifier during the different iterations. We denote by $V(y, x)$ the number of times that the label y was assigned in the example x during the previous iterations.

Then, we denote as *Vote Entropy* of an example x as:

$$VE(x) = - \sum_y \frac{V(y, x)}{iters} \log \frac{V(y, x)}{iters} \quad (3)$$

where *iters* refers to the number of iterations.

The examples to be labeled are chosen using equation (3), that is, examples with higher entropies are selected. As long as we add new examples during the iterations, the labeling of some examples will change as, new information will be given to the classifier. The strategy chooses the examples for which the classifier changes its decision more often during the iterations. We have to note, that during the first 2-3 iterations, we do not have enough information in order to choose the best examples to label, but very quickly the active learner manage to identify these examples. The intuition behind this model is that examples which tend to change labels are those for which the classifier seems more undecided.

Algorithm 2 gives us the general framework under which the above active learning methods can be combined with the semi-supervised variant of the PLSA model.

5 Experiments

In our experiments we used four different datasets: two collections from the CMU World Wide Knowledge Base project - **WebKB**¹ [4] and **20Newsgroups**², the widely used text collection of **Reuters** (**Reuters** – 21578)³ and a real-world dataset from Xerox. As mentioned before, we are concentrated in document classification; nevertheless, the algorithms described in the previous sections can be also used for different applications in which there is a relation of co-occurrence between objects and variables such as image classification.

These three datasets were pre-processed by removing the email tags and other numeric terms, discarding the tokens which appear in less than 5 documents and removing a total of 608 stopwords from the **CACM** stoplist⁴. No other form of pre-processing (stemming, multi-word recognition etc.) was used on the documents. Table 1 summarizes the characteristics of these datasets.

¹ <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/>

² <http://people.csail.mit.edu/jrennie/20Newsgroups/>

³ <http://www.daviddlewis.com/resources/testcollections/reuters21578/>

⁴ http://ir.dcs.gla.ac.uk/resources/test_collections/cacm/

Table 1. Characteristics of the datasets

Dataset	20Newsgroups	WebKB	Reuters
Collection size	20000	4196	4381
# of classes, K	20	4	7
Vocabulary size, $ \mathcal{W} $	38300	9400	4749
Training set size, $ X_l \cup X_u $	16000	3257	3504
Test set size	4000	839	876

Except the previous datasets which are widely used for evaluation of different classification algorithms in the Machine Learning community, we used a real world dataset (called **XLS**) which comes from a Xerox Business Group (XLS). This dataset is constituted of 20000 documents in the training set and 34770 in the test set. The documents consist of approximately 40% emails, 20% Microsoft Word documents; 20% Microsoft Excel documents, 10% Microsoft Power point documents and 10% PDF and miscellaneous documents. We want to classify the documents as *Responsive* and *Non-Responsive* to a particular given case.

Evaluation Measures. In order to evaluate the performance of the models, we used the microaverage F-score measure.

For each classifier, $\mathcal{G}_f$, we first compute its microaverage precision P and recall R by summing over all the individual decisions it made on the test set:

$$R(\mathcal{G}_f) = \frac{\sum_{k=1}^K \theta(k, \mathcal{G}_f)}{\sum_{k=1}^K (\theta(k, \mathcal{G}_f) + \psi(k, \mathcal{G}_f))}$$

$$P(\mathcal{G}_f) = \frac{\sum_{k=1}^K \theta(k, \mathcal{G}_f)}{\sum_{k=1}^K (\theta(k, \mathcal{G}_f) + \phi(k, \mathcal{G}_f))}$$

Where, $\theta(k, \mathcal{G}_f)$, $\phi(k, \mathcal{G}_f)$ and $\psi(k, \mathcal{G}_f)$ respectively denote the true positive, false positive and false negative documents in class k found by $\mathcal{G}_f$, and K denotes the number of classes. The F-score measure is then defined as [14]:

$$F(\mathcal{G}_f) = \frac{2P(\mathcal{G}_f)R(\mathcal{G}_f)}{P(\mathcal{G}_f) + R(\mathcal{G}_f)}$$

5.1 Results

We run experiments for all semi-supervised variants, for both active learning techniques, and for all four datasets. In our experiments, we label one example in each iteration and 100 iterations are performed for **WebKB**, **Reuters** and 150 for **20Newsgroups** dataset. For the **XLS** dataset we label 2 examples in each iteration, and we perform 100 iterations (as the dataset is bigger than the other three we need more data for achieving a good performance). For the *Margin Method*, it is not wise to choose 2 examples that are next to each other, as they cannot gives us more information that each of them does. As a result, we chose

Fig. 1. F-Score (y-axis) versus, the number of labeled examples in the training set $|D_l|$, (x-axis) graphs for the combination of the two ssPLSA algorithms with active learning on Reuters, WebKB and 20Newsgroups datasets

the two examples with the biggest class-entropy but, in addition, with different assigned labels.

In order to evaluate the performance of the active learning methods, we also run experiments for the combination of the semi-supervised algorithms with a random selection method, where in each iteration the documents to be labeled are chosen randomly.

As we can notice from the figure 1 the use of active learning helps, in comparison with the random query for all four datasets. The performance of the two different active learning techniques are comparable, and their difference is not statistically significant. Nevertheless, they clearly outperform the random method, especially when very few labeled data are available.

For the XLS dataset in particular, as we can notice, active learning helps, comparing to the random method, although the gain is less than the other three datasets. As in the previous case, the two active learning methods give similar results.

6 Conclusions

In this work, a variant of the semi-supervised PLSA algorithm has been combined with two active learning techniques. Experiments on four different datasets validate a consistent significant increase in performance. The evaluation we performed has shown that this combination can further increase classifier's

performance. Using active learning we manage to chose our training labeled set carefully, using the most informative examples. Working this way, we can achieve a better performance using less labeled examples.

This work was focused on the PLSA model. Nevertheless, this does not mean that the developed models can exclusively used with it. On the contrary, the proposed techniques are very easily applicable to different aspect models. Another possible extension is the use of different active learning techniques. Also, the combination of more than one active learning technique could be considered.

Acknowledgment

This work was supported in part by the IST Program of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

References

1. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. *Journal of Machine Learning Research* (2003)
2. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: *Proceedings of the Conference on Computational Learning Theory (COLT)*, pp. 92–100 (1998)
3. Campbell, C., Cristianini, N., Smola, A.J.: Query learning with large margin classifiers. In: *Proceedings of the 17th International Conference on Machine Learning (ICML)*, San Francisco, CA, USA, pp. 111–118 (2000)
4. Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., Slattery, S.: Learning to extract symbolic knowledge from the World Wide Web. In: *Proceedings of the 15th Conference of the American Association for Artificial Intelligence*, Madison, US, pp. 509–516 (1998)
5. Dagan, I., Engelson, S.P.: Committee-based sampling for training probabilistic classifiers. In: *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 150–157 (1995)
6. Davy, M., Luz, S.: Active learning with history-based query selection for text categorisation. In: Amati, G., Carpineto, C., Romano, G. (eds.) *ECiR 2007*. LNCS, vol. 4425, pp. 695–698. Springer, Heidelberg (2007)
7. Dönmez, P., Carbonell, J.G., Bennett, P.N.: Dual strategy active learning. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) *ECML 2007*. LNCS (LNAI), vol. 4701, pp. 116–127. Springer, Heidelberg (2007)
8. Freund, Y., Seung, H., Shamir, E., Tishby, N.: Selective sampling using the query by committee algorithm. *Machine Learning* 28(2-3), 133–168 (1997)
9. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning* 42(1-2), 177–196 (2001)
10. Iyengar, V., Apte, C., Zhang, T.: Active learning using adaptive resampling. In: *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, pp. 92–98 (2000)

11. Krithara, A., Amini, M.-R., Renders, J.-M., Goutte, C.: Semi-supervised document classification with a mislabeling error model. In: European Conference on Information Retrieval (ECIR), Glasgow, Scotland (2008)
12. Krithara, A., Goutte, C., Amini, M.-R., Renders, J.-M.: Reducing the annotation burden in text classification. In: 1st International Conference on Multidisciplinary Information Sciences and Technologies (InSCiT), Merida, Spain (2006)
13. Lewis, D., Gale, W.A.: A sequential algorithm for training text classifiers. In: Proceedings of the 17th International Conference on Research and Development in Information Retrieval (SIGIR), Dublin, pp. 3–12 (1994)
14. Lewis, D., Ringuette, M.: A comparison of two learning algorithms for text categorization. In: Proceedings of the Symposium on Document Analysis and Information Retrieval (SDAIR), pp. 81–93 (1994)
15. McCallum, A., Nigam, K.: Employing EM and pool-based active learning for text classification. In: Proceedings of the 15th International Conference on Machine Learning (ICML), pp. 350–358 (1998)
16. Muslea, I., Minton, S., Knoblock, C.: Active + Semi-supervised Learning = Robust Multi-View Learning. In: Proceedings of the 19th International Conference on Machine Learning (ICML), pp. 435–442 (2002)
17. Muslea, I., Minton, S., Knoblock, C.A.: Selective sampling with redundant views. In: Proceedings of the 17th National Conference on Artificial Intelligence (AAAI), pp. 621–626 (2000)
18. Probst, K., Ghani, R.: Towards 'interactive' active learning in multi-view feature sets for information extraction. In: Proceedings of European Conference on Machine Learning (ECML), pp. 683–690 (2007)
19. Roy, N., McCallum, A.: Toward optimal active learning through sampling estimation of error reduction. In: Proceedings of the 18th International Conference on Machine Learning (ICML), San Francisco, CA, pp. 441–448 (2001)
20. Seung, H.S., Oppor, M., Sompolinsky, H.: Query by committee. *Computational Learning Theory*, 287–294 (1992)
21. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. In: Proceedings of 17th International Conference on Machine Learning (ICML), Stanford, US, pp. 999–1006 (2000)
22. Tr, G., Hakkani-Tr, D., Schapire, R.: Combining active and semi-supervised learning for spoken language understanding. *Speech Communication* 45(2), 171–186 (2005)
23. Zhou, Z.-H., Chen, K.-J., Dai, H.-B.: Enhancing relevance feedback in image retrieval using unlabeled data. *ACM Transactions on Information Systems* 24(2), 219–244 (2006)
24. Zhu, X., Lafferty, J., Ghahramani, Z.: Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In: Proceedings of the 20th International Conference on Machine Learning (2003)