Graph-Based Semi-Supervised Learning for Semantic Role Diffusion

M. G. Carneiro[1], L. Zhao[2], J. L. G. Rosa[2]

[1] Universidade Federal de Uberlândia, Brazil
mgcarneiro@ufu.br
[2] Universidade de São Paulo, Brazil
zhao@usp.br, joaoluis@icmc.usp.br

Abstract. Semi-supervised learning (SSL) has recently attracted a considerable amount of research in machine learning. Among many categories of SSL techniques, graph-based methods are characterized by their ability to identify classes of arbitrary distributions. In this article we investigate the application of these methods for Semantic Role Labeling (SRL) which is the task of automatically identifying and classifying arguments with roles that indicate semantic relationship between an event and its participants. Such roles have great potential to improve a wide range of tasks, such as information extraction and machine translation. Experiments on a Brazilian Portuguese corpus named PropBank-br, which was built with text from Brazilian newspapers, were performed varying the number of labeled points, graph construction and label diffusion methods. The results show that the combination between symmetric k-nearest neighbors graph and local and global consistency method is a promising choice to semantic role diffusion on PropBank-br.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications; I.2.6 [**Artificial Intelligence**]: Learning; I.2.7 [**Artificial Intelligence**]: Natural Language Processing

Keywords: Graph-based SSL, Label Diffusion, PropBank-br, Semantic Role Labeling, Semi-supervised learning

1. INTRODUCTION

Semi-supervised learning (SSL) considers the general problem of learning from labeled and unlabeled data. It has turned out to be a new topic of machine learning research that has recently attracted a considerable amount of research [Chapelle et al. 2006; Zhu 2008]. While unlabeled data is far easier to obtain, the labeling process is often expensive, time consuming and requires the efforts of human annotators, who must often be quite skilled. Among many categories of SSL techniques which include generative models and low-density separation models, graph-based methods have as main advantage their ability to identify classes of arbitrary distributions [Silva and Zhao 2012]. See [Zhu et al. 2003; Zhou et al. 2004; Jebara et al. 2009; Ozaki et al. 2011; de Sousa et al. 2013] for more details about graph-based SSL.

In this article, we investigate graph-based SSL methods for Semantic Role Labeling (SRL) which is the task of automatically identifying and classifying the arguments of a predicate with roles. Such roles indicate meaningful relations among the arguments, as who did what to whom, where, when and how [Palmer et al. 2010]. Motivated by its potential to improve applications in a wide range of Natural Language Processing (NLP) tasks, such as information extraction, question answering, plagiarism detection, and so on, SRL has received much attention in the last years. In addition, many lexical resources, such as PropBank and FrameNet, have been built to allow the development of efficient semantic role labelers. Under the PropBank annotation framework each predicate is associated with

Authors thank the financial support given by the São Paulo State Research Foundation - FAPESP (grants numbers 2012/07926-3, 2011/50151-0 and 2013/07375-0). Authors also acknowledge support from CAPES and CNPq.

a set of core roles (named A0, A1, A2, and so on) which interpretation is specific to that predicate, and a set of adjunct roles (e.g., location, manner or time) which interpretation is common across predicates. Following we present an example about the SRL task through the sentences 1, 2 and 3. In 2, the arguments are identified, and the argument classification is showed in 3. The former aims to identify groups of words in a sentence that represent semantic arguments, and the latter aims to assign specific labels to the identified groups.

1. *Seymour Cray can do it again.*[1]
2. [*Seymour Cray*$_{arg}$] [*can*$_{arg}$] do [*it*$_{arg}$] [*again*$_{arg}$].
3. [*Seymour Cray*$_{A0}$] [*can*$_{mod}$] do [*it*$_{A1}$] [*again*$_{tmp}$].

To be specific, this work focus on the argument classification task using the PropBank-br [Duran and AluÃŋsio 2012] which is a Brazilian Portuguese corpus built with text from Brazilian newspapers that follows the PropBank style. An important motivation of this choice is related to the scarcity of annotated data in PropBank-br which size is about one seventh of the original PropBank [Fonseca and Rosa 2013]. This represents a difficulty scenario for machine learning where graph-based SSL methods can be evaluated through very arbitrary and unbalanced distributions. As the core roles are verb-dependant and in order to have a bigger number of unlabeled points, we perform experiments using three of the most frequent verbs in the PropBank-br, "dar" (to give), "fazer" (to do) and "dizer" (to say), which are evaluated using many graph construction methods and label diffusion strategies, and considering different number of labeled points. For graph construction, two widely used graph construction methods: symmetric k-nearest neighbors (SkNN) and mutual k-nearest neighbors (MkNN) are considered and they are also combined with the minimum spanning tree (MST) forming other two methods (SkNN+MST and MkNN+MST). For label diffusion, the gaussian field and harmonic function (GFHF)[Zhu et al. 2003] and local and global consistency (LGC) [Zhou et al. 2004] are analyzed.

The remainder of the article is organized as follows. Sect. 2 briefly presents some related works; Sect. 3 describes the framework adopted in this work. Computer simulations are presented in Sect. 4; and Sect. 5 concludes the article.

2. RELATED WORK

The objective of this work is simple. It aims to investigate graph-based SSL methods for SRL task in order to explore particular characteristics of the PropBank.br such as scarcity of labeled data and unbalanced class distributions.

Most part of the techniques developed for SRL lies on the supervised category [Gildea and Jurafsky 2002; Pradhan et al. 2008] where large (and expensive) amount of human annotated data are used to train classifiers. Although PropBank-br annotated data is scarse, some works explored it in the supervised context [Alva-Manchego and Rosa 2012; Fonseca and Rosa 2013; Hartmann et al. 2016; Carneiro et al. 2016].

SRL literature contains few works about unsupervised and semi-supervised learning. Unsupervised works such as presented in [Lang and Lapata 2011] refer to the task as semantic role induction and the objective is to cluster argument instances of each verb. SSL works include the investigation of semi-supervised and "semi-unsupervised" approaches. For instance, bootstrapping approaches such as self-training and co-training methods are proposed in [He and Gildea 2006] while a semi-unsupervised approach which employs a small number of labeled data to build an informed prior distribution over an unsupervised method is presented in [Titov and Klementiev 2012]. In addition, approaches

[1] `http://verbs.colorado.edu/propbank/framesets-english/do-v.html`

that increase the manually annotated instances with unlabeled instances which roles were infered through projection [Fürstenau and Lapata 2012] as well as the reduction of lexical features sparsity by exploiting word representation [Deschacht and Moens 2009] have been proposed.

3. GRAPH-BASED SEMI-SUPERVISED LEARNING

Problem definition. Given a set of arguments $\mathbf{X} = \{a_1, \ldots, a_l, a_{l+1}, \ldots, a_n\}$ and a set of semantic roles $\mathcal{L} = 1, \ldots, c$, the first l arguments are labeled $\{y_1, \ldots, y_l\} \in \mathcal{L}$ and the remaining arguments $(u = n - l)$ are unlabeled. Tipically, $l \ll u$, i.e., the great majority of arguments does not possess labels. The goal is to predict the semantic roles of the unlabeled arguments.

General framework. Following we present the general framework used in our investigation:

(1) Build an undirected graph $\mathcal{G}$ from $\mathbf{X}$;
(2) Generate a weighted matrix $\mathbf{W}$ from a similarity measure $\mathcal{K}$ and $\mathcal{G}$;
(3) Compute the graph Laplacian matrix $\mathbf{L}$ from $\mathbf{W}$;
(4) Label the unlabeled points from the output matrix $\mathbf{F}$ obtained by using a diffusion strategy.

3.1 Graph Construction

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where each node $v_i \in V$ represents an argument $a_i \in \mathbf{X}$. Let $\mathbf{S}$ be a distance matrix in which $\mathbf{S}_{ij} = \delta(a_i, a_j)$ and $k\mathrm{NN}_i$ be the set of k nearest neighbors of a_i, the adjacency matrix $\mathbf{A}$ of a kNN-graph is obtained as follows:

$$\mathbf{A}_{ij} = \begin{cases} 1, & \text{if } a_j \in k\mathrm{NN}_i \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

As the kNN-graph may not be symetric, two strategies are commonly used to symmetrize it: symmetric kNN and mutual kNN. The symmetric kNN (SkNN) is obtained as follows:

$$\mathbf{A} = max(\mathbf{A}, \mathbf{A}^T) \ , \tag{2}$$

and the mutual kNN (MkNN) is obtained as follows:

$$\mathbf{A} = min(\mathbf{A}, \mathbf{A}^T) \ . \tag{3}$$

We also investigate the combination of both graph construction methods to the minimum spanning tree graph (MST) as a way to rigorously ensure the graph-based SSL assumption that each unlabeled point is on a connected subgraph in which there exists at least one labeled point. Let $\mathcal{G}$ denote a fully connected graph whose weights of the edges are given by a distance matrix $\mathbf{S}$. The minimum spanning tree (MST) of $\mathcal{G}$ is the subset of edges $\mathcal{E}' \subset \mathcal{E}$ which connect all the nodes in $\mathcal{V}$ minimizing the following quantity: $\sum_{i,j} \mathbf{S}_{ij}$.

3.2 Weighted Matrix Generation and Graph Laplacian

Despite the possibility to generate a fully connected weighted matrix, the graph construction step is important to promote sparsification in $\mathbf{W}$ which considerably reduces the computational complexity and usually improves the performance. Thus, a symmetric weighted matrix $W \subset \mathbb{R}^{n \times n}$ is given by:

$$\mathbf{W}_{ij} = \mathbf{A}_{ij}\mathcal{K}(a_i, a_j), \tag{4}$$

where $\mathcal{K}(a_i, a_j)$ denotes a similarity function. The Gaussian or RBF kernel has been used in this article:

$$\mathcal{K}(a_i, a_j) = exp\left(-\frac{\delta(a_i, a_j)}{2\sigma^2}\right), \tag{5}$$

where $\delta(\cdot)$ is a vector-based distance function such as the l_2-norm $\|a_i - a_j\|^2$, and σ is the kernel bandwidth parameter.

From the weighted matrix $\mathbf{W}$, the normalized Laplacian $\mathbf{L}$ is obtained:

$$\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}, \tag{6}$$

where $\mathbf{I}_n$ is the identity matrix and $\mathbf{D}$ is the diagonal matrix which contains the vertices degree. As the literature suggest the normalized Laplacian may lead better results in comparison with the combinatorial Laplacian ($\mathbf{\Delta} = \mathbf{D} - \mathbf{W}$), we used $\mathbf{L}$ in the formulation of the label diffusion algorithms which are presented in next sub-section.

3.3 Label Diffusion

The label diffusion process is conducted by the following graph-based techniques: gaussian field and harmonic function (GFHF) [Zhu et al. 2003] and local and global consistency (LGC) [Zhou et al. 2004]. Given a graph Laplacian $\mathbf{L}$ and let $\mathbf{Y} \in \mathbb{B}^{n \times c}$ be a label matrix in which $\mathbf{Y}_{ij} = 1$ if and only if a_i has label $y_i = j$, GFHF and LGC generate the output matrix $\mathbf{F}$ by label diffusion as follows:

Gaussian Field and Harmonic Function GFHF obtains the output matrix $\mathbf{F}$ as follows:

$$\mathbf{F}_{\mathcal{U}} = -\mathbf{L}_{\mathcal{U}\mathcal{U}}^{-1}\mathbf{L}_{\mathcal{U}\mathcal{L}}\mathbf{Y}_{\mathcal{L}} \ . \tag{7}$$

Local and Global Consistency LGC obtains the output matrix $\mathbf{F}$ as follows:

$$\mathbf{F}(t+1) = \alpha\mathbf{L}\mathbf{F}(t) + (1-\alpha)\mathbf{Y} \ , \tag{8}$$

where α defines the relative amount of the information from its neighbors and its initial label.

4. COMPUTER SIMULATIONS

This section provides experimental results using the general framework described in the previous section. The objective is to evaluate the performance of the label diffusion techniques on the SRL task. In the study, each simulation was performed in a transductive setting using different numbers of labeled and unlabeled points. The number of labeled points l is dependant of the number of classes c in the data set, i.e., $l = \eta c$ with $\eta \in \{1, 2, 3\}$. For each labeled set size l tested, we perform 30 trials. In each trial we randomly sample labeled data from the entire data set ensuring there is at least one labeled point per class, and use the rest of instances as unlabeled data. The error rate averaged over the trials is used to evaluate the quality of the semantic role diffusion process.

4.1 Datasets

In this article we present results for SRL task using the PropBank.br. For the experiment, we select all sentences related to the predicate “to give”, “to do” and “to say”, which are among the most frequent verbs in the corpus, and extract the attributes of each argument by using a set of features from the literature[2] [Gildea and Jurafsky 2002; Alva-Manchego and Rosa 2012; Pradhan et al. 2008]. As a pre-processing step, argument classes smaller than ten instances were excluded. Table I presents a brief description of the data sets obtained, named here “PBbr-give”, “PBbr-do” and “PBbr-say”, in terms of number of instances/arguments and classes. In the simulations, each argument was mapped as a vertex into an underlying network. As a data preparation, each instance attribute vector was normalized to have a magnitude of one and the euclidean distance was used in all simulations as

[2]The features used were: FirstForm+FirstPostag, FirstLemma, Head, HeadLemma, TopSequence, PostagSequence, PredLemma+PhraseType, LastForm+LastPostag, PredLemma+Path, FirstPostag, LeftHead, RightHead, VoicePosition, LeftHeadPostag, RightPhrase, and PredLemma.

the distance measurement. In order to avoid the dimensionality curse problem, we run principal component analysis (PCA) to reduce the dimensionality of the data to 100 features.

Table I: Brief description of the PropBank-br data sets.

	PBbr-give	PBbr-do	PBbr-say
#Instances / #Classes	148 / 3	397 / 8	506 / 5

4.2 Parameters and Baseline

The following parameters are employed in the computer simulations presented here: the value of k in the kNN-graph is optimized over the set $k \in \{1, 2, \ldots, 60\}$; the kernel bandwidth σ in the Gaussian Kernel $\mathcal{K}$ is defined as suggested in [Jebara et al. 2009] by $\sigma = \bar{d}_k/3$ where $\bar{d}_k$ is the average distance between each sample and its k-th nearest neighbor; and the parameter α in LGC is chosen at range $\{0.001, 0.01, 0.05, 0.1, 0.2, 0.5, 0.8, 1.0\}$. As in related works in literature, 1NN classifier is used for comparison purposes as a baseline.

4.3 Results and Discussion

This subsection provides results obtained in the simulations which are organized in three major experiments according to the number of labeled points $\eta \in \{1, 2, 3\}$. Note that $\eta = 1$ is the most relevant experiment to the addressed problem as it suggests approximately one sentence per verb is labeled.

The results of the three experiments are presented in Table II. In the first one ($\eta = 1$), which employs one labeled item per class, one can see: (i) LGC-SkNN presents the best performance on PBbr-give and PBbr-do data sets while GFHF-SkNN performs very well on PBbr-say; (ii) the baseline performance is very far from both GFHF and LGC when they are combined with SkNN and SkNN+MST graph construction methods. In the second and third one ($\eta = \{2, 3\}$), one can observe: (i) LGC-* (the four graph construction methods have very similar performance) presents the better performance on PBbr-give and LGC-SkNN on PBbr-do, and *-SkNN+MST performs very well on PBbr-say; (ii) MkNN graph construction has the worst performance on all data sets; (iii) by increasing the number of labeled points ($\eta > 1$), the baseline has its performance considerably improved on the data sets, but it is also outperformed. In order to analyze statistically the results obtained in the simulations, the Wilcoxon Signed Ranks test is adopted with $\alpha = 0.1$. The tests reveal LGC-SkNN and LGC-SkNN+MST provide the better results when comparing each two methods over all data sets.

Now we move on to analyze the general performance of both label diffusion algorithms GFHF and LGC in function of the graph construction methods and the variation of their parameter k. As LGC has also the parameter α, we set α according to the best performance obtained by the technique. Again we consider our division of the computer simulations in three experiments according to the number of labeled points. Figs. 1, 2 and 3 show the average error rates of the techniques for $\eta = 1$ (one labeled point per class), $\eta = 2$ and $\eta = 3$, respectively. By analyzing the figures, we can see:

GFHF × LGC: Despite GFHF does not require parameter selection (σ is defined heuristically), its predictive performance is usually worse than LGC in the PropBank-br data sets considered here.

SkNN × MkNN: Independently of the number of labeled points l and the label diffusion algorithm used, MkNN present the worst general performance on PBbr-give as well as on PBbr-say data sets. SkNN graph construction method is also better than MkNN on PBbr-do data set, with exception in the case when there is only one labeled point per class ($\eta = 1$).

MST × not MST: For $\eta > 1$, the minimum spanning tree (MST) graph improves the general performance of both SkNN and MkNN on PBbr-give and PBbr-say data sets.

Table II: Comparative results in terms of average error rates and standard deviations (over thirty runs) using $\eta = 1$ (one labeled point per class), $\eta = 2$ and $\eta = 3$. Best result for each data set is in bold face.

$\eta = 1$		PBbr-give (3)	PBbr-do (8)	PBbr-say (5)
	1NN	45.31 (11.37)	63.52 (9.80)	35.41 (14.97)
	GFHF-SkNN	41.26 (11.20)	50.21 (4.50)	**29.41 (19.84)**
	GFHF-SkNN+MST	41.54 (9.96)	50.21 (4.50)	30.95 (15.03)
	GFHF-MkNN	43.72 (9.05)	50.75 (4.52)	34.55 (7.93)
	GFHF-MkNN+MST	43.84 (8.70)	50.34 (4.53)	34.42 (7.16)
	LGC-SkNN	**37.65 (11.20)**	**47.02 (2.90)**	30.31 (19.36)
	LGC-SkNN+MST	**37.65 (11.20)**	47.78 (3.79)	30.82 (16.26)
	LGC-MkNN	39.61 (9.69)	48.93 (3.03)	34.47 (7.96)
	LGC-MkNN+MST	39.72 (10.38)	47.13 (3.30)	33.59 (4.01)
$\eta = 2$		PBbr-give (6)	PBbr-do (16)	PBbr-say (10)
	1NN	32.98 (8.96)	47.23 (3.77)	22.65 (13.15)
	GFHF-SkNN	31.78 (10.15)	44.28 (4.84)	23.47 (10.71)
	GFHF-SkNN+MST	31.73 (9.58)	45.74 (4.09)	**21.46 (12.36)**
	GFHF-MkNN	32.91 (9.21)	46.02 (4.28)	26.16 (7.77)
	GFHF-MkNN+MST	32.14 (9.92)	45.82 (4.12)	23.56 (12.10)
	LGC-SkNN	30.96 (9.43)	**43.57 (4.18)**	22.91 (10.83)
	LGC-SkNN+MST	30.96 (9.43)	44.82 (4.73)	21.70 (14.49)
	LGC-MkNN	30.96 (9.52)	46.63 (3.24)	26.04 (7.81)
	LGC-MkNN+MST	**30.89 (9.36)**	44.98 (5.67)	22.00 (12.67)
$\eta = 3$	Alg.	PBbr-give (9)	PBbr-do (24)	PBbr-say (15)
	1NN	33.76 (9.51)	43.22 (6.26)	15.48 (6.57)
	GFHF-SkNN	29.50 (8.64)	42.48 (5.07)	17.05 (8.69)
	GFHF-SkNN+MST	29.50 (8.64)	43.13 (4.28)	15.80 (8.26)
	GFHF-MkNN	30.36 (8.71)	43.44 (4.24)	20.98 (7.79)
	GFHF-MkNN+MST	29.86 (8.73)	43.34 (4.28)	15.84 (8.37)
	LGC-SkNN	**28.94 (8.36)**	**41.38 (5.07)**	16.71 (8.69)
	LGC-SkNN+MST	**28.94 (8.36)**	44.03 (5.09)	**15.12 (7.35)**
	LGC-MkNN	29.18 (8.74)	45.29 (4.87)	19.81 (8.17)
	LGC-MkNN+MST	29.02 (8.91)	43.31 (5.48)	16.00 (7.92)

PBbr-do: The best average error rates obtained to PBbr-do usually have small values of k. We believe this is because the data set is very unbalanced with 2 of 8 classes pursuing more than 68% of the data points. In this sense, LGC have some trouble to decrease its average error rate as k increases while GFHF not. Such result deserves more investigation.

5. CONCLUSION

In this article we investigated the application of graph-based SSL methods for semantic role labeling on a Brazilian Portuguese corpus named PropBank-br. Four graph construction methods and two label diffusion functions were evaluated in computer simulations with different number of labeled points. The results show LGC-SkNN as the best combination. The analysis also reveal MST can improve the general performance of the techniques as the number of labeled points increase. Forthcoming works include comparisons with other SSL methods and the inclusion of more verbs in the experiments.

REFERENCES

Alva-Manchego, F. E. and Rosa, J. L. G. Semantic role labeling for brazilian portuguese: A benchmark. In *Advances in Artificial Intelligence – IBERAMIA 2012*. Springer, pp. 481–490, 2012.

Carneiro, M. G., Zhao, L., Cheng, R., and Jin, Y. Network structural optimization based on swarm intelligence for highlevel classification. In *IEEE International Joint Conference on Neural Networks*. pp. 3737–3744, 2016.

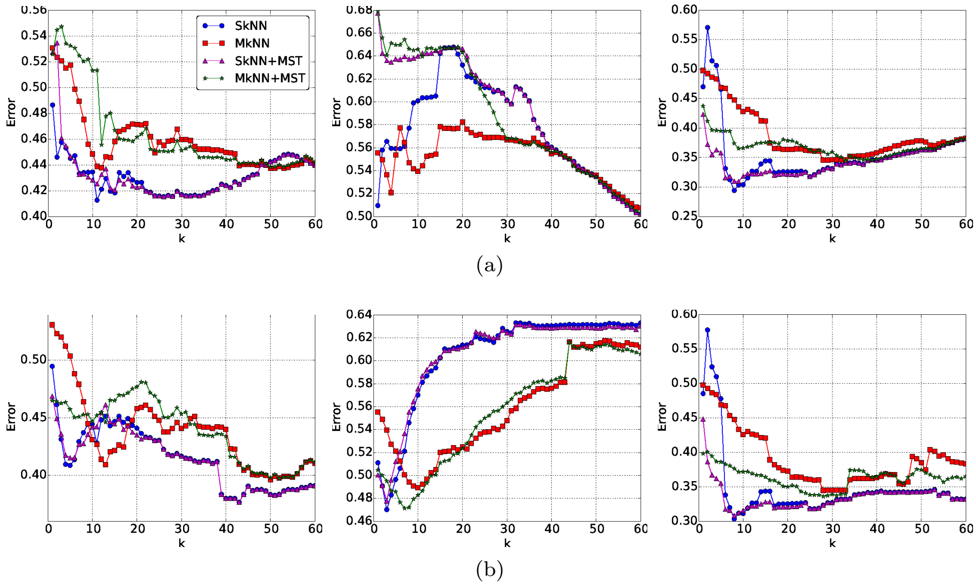


Fig. 1: Average error rates of (a) GFHF and (b) LGC on PBbr-give (left), PBbr-do (middle) and PBbr-say (right) data sets with one labeled point per class.

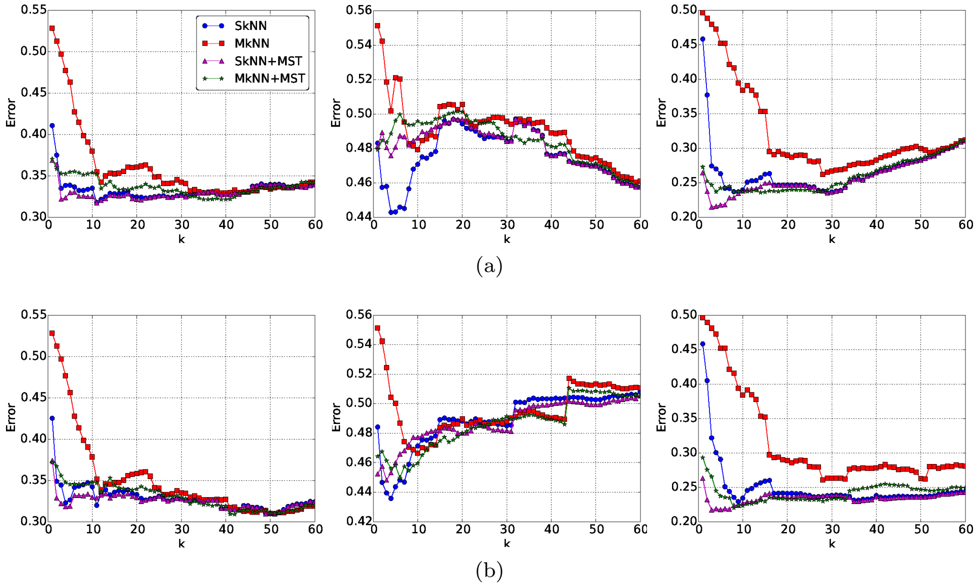


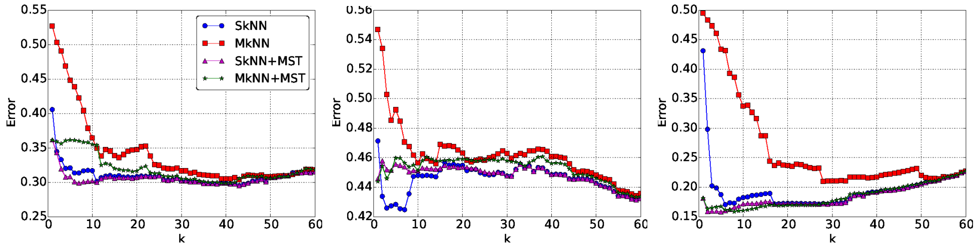
Fig. 2: Average error rates of (a) GFHF and (b) LGC on PBbr-give (left), PBbr-do (middle) and PBbr-say (right) data sets with the number of labeled points given by 6, 16 and 10, respectively.

Chapelle, O., Schölkopf, B., and Zien, A. *Semi-supervised learning.* MIT Press, 2006.

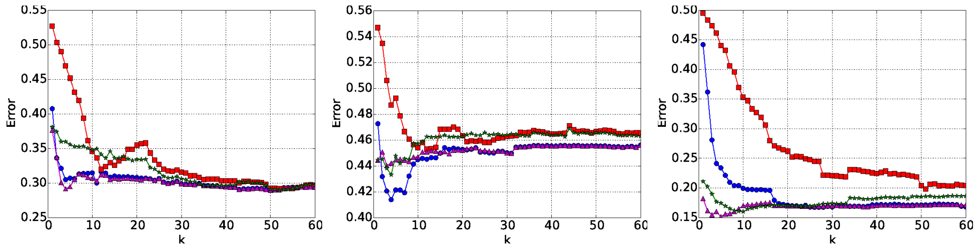
de Sousa, C. A. R., Rezende, S. O., and Batista, G. E. Influence of graph construction on semi-supervised learning. In *Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 160–175, 2013.

Deschacht, K. and Moens, M.-F. Semi-supervised semantic role labeling using the latent words language model. In *ACL Conference on Empirical Methods in Natural Language Processing*. pp. 21–29, 2009.

Duran, M. S. and AluÃDsio, S. M. Propbank-br: a brazilian treebank annotated with semantic role labels. In *International Conference on Language Resources and Evaluation*. pp. 1862–1867, 2012.



(a)



(b)

Fig. 3: Average error rates of (a) GFHF and (b) LGC on PBbr-give (left), PBbr-do (middle) and PBbr-say (right) data sets with the number of labeled points given by 9, 24 and 15, respectively.

Fonseca, E. R. and Rosa, J. L. G. A two-step convolutional neural network approach for semantic role labeling. In *IEEE International Joint Conference on Neural Networks*. pp. 2955–2961, 2013.

Fürstenau, H. and Lapata, M. Semi-supervised semantic role labeling via structural alignment. *Computational Linguistics* 38 (1): 135–171, 2012.

Gildea, D. and Jurafsky, D. Automatic labeling of semantic roles. *Computational Linguistics* 28 (3): 245–288, 2002.

Hartmann, N. S., Duran, M. S., and Aluísio, S. M. Automatic semantic role labeling on non-revised syntactic trees of journalistic texts. In *Computational Processing of the Portuguese Language*. pp. 202–212, 2016.

He, S. and Gildea, D. Self-training and co-training for semantic role labeling: Primary report. Tech. rep., 2006.

Jebara, T., Wang, J., and Chang, S.-F. Graph construction and b-matching for semi-supervised learning. In *International Conference on Machine Learning*. pp. 441–448, 2009.

Lang, J. and Lapata, M. Unsupervised semantic role induction via split-merge clustering. In *Annual Meeting of the Association for Computational Linguistics*. Vol. 1. pp. 1117–1126, 2011.

Ozaki, K., Shimbo, M., Komachi, M., and Matsumoto, Y. Using the mutual k-nearest neighbor graphs for semi-supervised classification of natural language data. In *ACL Conference on Computational Natural Language Learning*. pp. 154–162, 2011.

Palmer, M., Gildea, D., and Xue, N. *Semantic Role Labeling*. Morgan & Claypool Publishers, 2010.

Pradhan, S. S., Ward, W., and Martin, J. H. Towards robust semantic role labeling. *Computational Linguistics* 34 (2): 289–310, 2008.

Silva, T. C. and Zhao, L. Network-based stochastic semisupervised learning. *IEEE Transactions on Neural Networks and Learning Systems* 23 (3): 451–466, 2012.

Titov, I. and Klementiev, A. Semi-supervised semantic role labeling: Approaching from an unsupervised perspective. In *International Conference on Computational Linguistics*. pp. 2635–2652, 2012.

Zhou, D., Bousquet, O., Lal, T. N., Weston, J., and Schölkopf, B. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 2004.

Zhu, X. Semi-supervised learning literature survey. Tech. rep., 2008.

Zhu, X., Ghahramani, Z., and Lafferty, J. Semi-supervised learning using gaussian fields and harmonic functions. In *International Conference on Machine Learning*. pp. 912–919, 2003.