The authors present a novel approach for inducing unsupervised part-of-speech (POS) taggers for languages that have no labeled training data but have translated text in a resource-rich language. They use graph-based label propagation for cross-lingual knowledge transfer and use the projected labels as features in an unsupervised model. Their approach outperforms the state-of-the-art baseline and vanilla hidden Markov models induced with the Expectation Maximization algorithm across eight European languages. The paper discusses the challenges of unsupervised POS tagging and presents experimental results to validate the effectiveness of their approach.In this section, the authors describe their approach to building part-of-speech (POS) taggers for foreign languages using an English POS tagger and parallel text between the two languages. They propose a bilingual similarity graph construction, which makes use of similarity functions to establish connections between foreign language trigram types and English word types. The graph construction does not require any labeled data and uses a co-occurrence based similarity function to compute the edge weights between foreign language trigrams. This similarity function indicates how syntactically similar the middle words of the connected trigrams are. In addition, the authors use an unsupervised word alignment statistics-based similarity function to establish a soft correspondence between the two languages. To initialize the graph, they tag the English side of the parallel text using a supervised model and generate label distributions for the English vertices by aggregating the POS labels of the English tokens to types. They then use label propagation to transfer the labels to the peripheral foreign vertices first and then among all of the foreign vertices. The POS distributions over the foreign trigram types are used as features to learn a better unsupervised POS tagger. The authors show that their approach achieves significantly higher average POS tagging accuracy compared to previous unsupervised models and bridges the gap to fully supervised POS tagging performance. This section describes the graph vertices used in the bilingual setup of the study. The vertices are divided into two types: foreign language vertices (Vf) and English vertices (Ve). The foreign language vertices correspond to trigram types from the foreign language, while the English vertices correspond to word types. The graph is asymmetric, flowing from English to the foreign language, so the types of vertices used are different for each language. The foreign language vertices are extracted from a parallel corpus (De, Df) and an additional unlabeled monolingual foreign corpus (Î“f). The English vertices are labeled, so they don't need to be disambiguated by embedding them in trigrams. Additionally, the English vertices are not connected to each other, but only to the foreign language vertices. To compute edge weights between foreign trigram types, a monolingual similarity function is used. This function measures the co-occurrence statistics of nine different feature concepts, such as trigram + context, left context, right context, etc. For each trigram type, the function counts how many times it co-occurs with each feature concept and computes the point-wise mutual information (PMI) between them. The similarity between two trigram types is then calculated by summing the PMI values over common feature instantiations. For the English and foreign vertices, a bilingual similarity function is defined based on high-confidence word alignments. Word alignment techniques are used on the parallel corpus to align the English sentences and the foreign trigrams. This alignment information is then used to define the similarity function between the English and foreign vertices. Overall, the graph is constructed using these vertices and similarity functions, allowing for a graph-based approach to semi-supervised learning in a bilingual setup.The objective of the graph optimization is to minimize the function C(q). The function consists of two terms: the graph smoothness regularizer and the regularizer that encourages all type marginals to be uniform. The graph smoothness regularizer penalizes neighboring vertices that have different label distributions, while the regularizer encourages the label distributions to be uniform. To optimize the objective, an iterative update method is used. The update is formulated as q(m)i(y) = ri(y) if ui belongs to Vlf (the set of labeled foreign language vertices), and Î³i(y)/Îºi otherwise. Î³i(y) and Îºi are defined based on neighboring vertices and hyperparameters Î½ and U. The procedure is run for 10 iterations. After label propagation, tag probabilities are computed for foreign word types by marginalizing the POS tag distributions of foreign trigrams over the left and right context words. A threshold value Ï„ is used to eliminate labels with probabilities below the threshold. The resulting vector tx is used as features for the unsupervised foreign language POS tagger. The POS induction model is developed based on the feature-based HMM of Berg-Kirkpatrick et al. A first order Markov model is used to define the distribution of the sentence and state sequence. The emission distribution is replaced with a log-linear model that incorporates overlapping features of the observation. Overall, the objective is to optimize the graph by encouraging similar vertices to have similar label distributions and by regularizing the label distributions towards uniformity. The optimization is done iteratively using an update method, and the resulting tag probabilities are used for POS induction.The objective function used to train the model is defined as: L(Î˜) = N/Î£i=1 log(Î£z PÎ˜(X=x(i),Z=z(i))) - C||Î˜||^2 Where N is the number of training examples, Î˜ is the set of model parameters, x(i) is the i-th training example, z(i) is the corresponding state configuration, PÎ˜(X=x(i),Z=z(i)) is the probability of the example and state configuration under the model, and C is a regularization parameter. The model uses features to check if the word identity x contains digits or hyphens, if the first letter of x is uppercase, and suffix features up to length 3. These features are conjoined with the state z. The model is trained using the L-BFGS optimization algorithm, a quasi-Newton method. This method has been found to perform better than using a feature-enhanced modification of the Expectation-Maximization (EM) algorithm for English POS tagging. The model also incorporates a novel constraint feature, which incorporates information from a smoothed graph and prunes hidden states inconsistent with a thresholded vector tx. The constraint feature is defined as: ft(x,z) = log(tx(y)), if Î»(z) = y Where tx(y) is the thresholded vector element for the tag y, and Î» is a function that maps from the language-specific tagset F to the universal tagset C. The model is evaluated using monolingual treebanks and parallel text datasets for eight Indo-European languages: Danish, Dutch, German, Greek, Italian, Portuguese, Spanish, and Swedish. The universal POS tagset of Petrov et al. is used in the experiments. The goal of the experiments is to apply the techniques to languages for which no labeled resources are available, so the same hyperparameters are used for all language pairs and no language-specific tuning is performed.The given text describes a study on part-of-speech (POS) tagging accuracy in various languages. The study focuses on a tagset consisting of 12 categories, which cover the most frequent POS in all the languages studied. The researchers provide a mapping from the language-specific POS tags in the foreign treebank to the universal POS tags. The experiment includes three baselines and two oracles in addition to two variants of a graph-based approach. The baselines include an EM-HMM, a Feature-HMM, and a Projection model. The graph-based approach includes a version with label propagation and a version without label propagation. The oracles involve using tagging dictionaries extracted from the treebanks and a supervised model trained on the original treebanks. The experimental setup involves setting hyperparameters for the models, such as the regularization constant and the number of iterations for training. The results show that the graph-based approach with label propagation outperforms the baselines and oracles in terms of POS tagging accuracy. The researchers conclude that their approach effectively utilizes bilingual information and improves the accuracy of POS tagging in different languages.