ASR Hypothesis Reranking using Prior-informed Restricted Boltzmann Machine

Yukun Ma[1,2], Erik Cambria[1], Benjamin Bigot[2]

[1]School of Computer Science and Engineering
[2]Rolls-Royce@NTU Corporate Lab
Nanyang Technological University, Singapore
mayu0010@e.ntu.edu.sg, cambria@ntu.edu.sg, bbigot@ntu.edu.sg

Abstract. Discriminative language models (DLMs) have been widely used for reranking competing hypotheses produced by an Automatic Speech Recognition (ASR) system. While existing DLMs suffer from limited generalization power, we propose a novel DLM based on a discriminatively trained Restricted Boltzmann Machine (RBM). The hidden layer of the RBM improves generalization and allows for employing additional prior knowledge, including pre-trained parameters and entity-related prior. Our approach outperforms the single-layer-perceptron (SLP) reranking model, and fusing our approach with SLP achieves up to 1.3% absolute Word Error Rate (WER) reduction and a relative 180% improvement in terms of WER reduction over the SLP reranker. In particular, it shows that the proposed prior informed RBM reranker achieves largest ASR error reduction (3.1% absolute WER) on content words.

1 Introduction

Reranking models have been shown effective for reducing errors in a variety of Natural Language Processing tasks such as Named Entity Recognition [1, 2], Syntactic Parsing [3, 4] and Statistical Machine Translation [5]. A reranking model typically treat the baseline system as a black box and is trained to rank the competing hypotheses based on more complex or global information.

In Automatic Speech Recognition (ASR), discriminative language model (DLM) was first introduced by Roark et al. [6] for reranking ASR hypotheses. They adopt a single perceptron to modify the confidence scores of hypotheses generated by a baseline ASR system. By using only n-gram features, their reranking model was shown capable of reducing the Word Error Rate (WER) of an ASR system. His work is followed by several variants with a variety of feature choices such as syntactic features [7, 8], which try to capture correlation between simple features on the feature level. However, existing DLMs still suffer from poor generalization power and are vulnerable to shortage of training data, because most of them rely on linear or log linear models that fail to take into consideration the correlation of input features on the model level. Apart from feature engineering, using hidden variables encoding semantic information helps improving the generalization power.

Koo et al. [4] proposes a hidden-variable model to rerank syntactic parsing trees. By linking input features to hidden states corresponding to word senses or classes, they achieves improved accuracy over a linear baseline. Inspired by the success of Koo et al., we propose to use the computational structure of Restricted Boltzmann Machine (RBM) [9] for the task of ASR hypotheses reranking. RBM is a neural network composed of one hidden layer and one input layer. The hidden layer of RBM has been shown capable of capturing high-order correlation and semantic information in the context of language modeling [10, 11]. These approaches model the probability of a fixed length of word sequences, i.e., N-grams, using only local information, and are trained with a generative objective function. However, RBM cannot be directly used for ASR reranking due to its generative training manner.

We propose two modifications to train RBM in a more task-specific way. We modify the energy function of RBM to incorporate the ASR confidence score, which has been proved critical for reranking by previous DLMs [6–8]. We then propose a novel discriminative objective function for training RBM with N-best lists of ASR hypotheses. Our method differs from existing RBM-based language models [10, 11] in two major aspects. Firstly, the proposed RBM reranker is trained discriminatively. Secondly, RBM in our method represents sentences of variable length as global feature vectors. Another attractive property of RBM is that the computational structure is flexible enough to incorporate various sources of prior knowledge [12]. As function words have little meanings and are less important for language understanding [13], we decide to focus more on content words, e.g., named entities. We hence further integrate to our model two types of prior knowledge: named entity related prior and a pre-trained hidden layer.

To our knowledge, this paper is the first to consider using hidden layer and prior knowledge in the context of ASR hypotheses reranking. The remainder of this paper is structured as follows: Section 2 describes in detail the proposed work; Section 3 shows the empirical results as well as analyses; finally, Section 4 concludes this paper and discusses about future work.

2 Training RBM for ASR Reranking

2.1 Restricted Boltzmann Machine

A Restricted Boltzmann Machine [9] (see Figure 1) is a neural network composed of : one n-dimension input feature layer $\phi(\mathbf{t}) = [\phi_1(t), \phi_2(t), \cdots, \phi_n(t)]$, which is a global feature vector extracted for a raw input t, and one d-dimension binary hidden layer $\mathbf{h} = [h_1, h_2, \cdots, h_d]$. The joint probability $P_{\mathrm{RBM}}(t, h)$ of hidden variables and raw input is defined as

$$P_{\mathrm{RBM}}(t,h) = \frac{e^{-E_{\mathrm{RBM}}(t,h)}}{\sum_{t,h} e^{-E_{\mathrm{RBM}}(t,h)}}$$
$$E_{\mathrm{RBM}}(t,h) = -\phi(t)^T W h - b^T \phi(t) - c^T h,$$

where $W \in R^{n \times d}$ is the matrix specifying the weights of connections between hidden and input layer, and $b \in R^n$ and $c \in R^d$ are the bias vectors of the two layers. $E_{\text{RBM}}(t, h)$ is called the energy function of RBM. The probability of a raw input t is then defined as the marginal probability of t

$$P_{\text{RBM}}(t) = \sum_h P_{\text{RBM}}(t, h)$$

, and the training objective is to maximize the log likelihood of training data D

$$\sum_{t \in D} \ln P_{\text{RBM}}(t)$$

2.2 Maximum Margin Training for RBM-based Reranker

The goal of generative training of RBM is to learn a probability distribution, which is not necessary for choosing correct ASR hypotheses. Instead, the discriminative training allows the model to explicitly select ASR hypotheses containing fewer errors. In this section, we describe our discriminatively trained RBM-bsed reranking model, denoted as **dRBM**.

Before introducing the training objective function, we first introduce the energy function of RBM model. ASR posterior probabilities produced by the baseline ASR system have been shown useful for reranking in previous works on DLM [6–8]. We hence add ASR posterior to the energy function of RBM. The modified energy function is expressed as

$$\begin{aligned} E_{\text{dRBM}}(t, h) &= E_{\text{RBM}}(t, h) + E_{asr}(t) \\ E_{asr}(t) &= -w_0 \ln(P(t|a)), \end{aligned}$$

where $P(t|a)$ is the posterior probability of a given ASR hypothesis t given the acoustic input a, and w_0 is the weight of ASR confidence score fixed during training. We represent each hypothesis as a global feature vector $\phi(t)$ using a predefined set of feature functions. In this paper, we mainly consider using unigram features, yet using more complicated features does not need to change the model.

Inspired by the maximum margin training for Bayesian Networks [14], we adopt a discriminative objective function L using likelihood ratio,

$$L = \frac{1}{|D|} \sum_{a \in D} \sum_{t' \in \text{GEN}(a)} \max(1 - \ln \frac{P_{\text{dRBM}}(\hat{t})}{P_{\text{dRBM}}(t')}, 0),$$

where D is the training set for the discriminative training of RBM and $|D|$ denotes the number of utterances in training set. $\text{GEN}(a)$ refers to the list of N-best hypotheses generated by the baseline ASR system for the acoustic input a, while $\hat{t}$ is the oracle-best in the N-best list of t. Intuitively, the learning process finds the parameter setting maximizing the margin between the oracle-best hypotheses and other hypotheses in the N-best list. The subgradient of the objective function is

$$\frac{\partial L}{\partial \theta} = \sum_{a \in D} \sum_{t' \in \mathrm{GEN}(a)} \mathbb{I}(\mathcal{F}(\hat{t}) - \mathcal{F}(t') < 1)(\frac{\partial \mathcal{F}(t')}{\partial \theta} - \frac{\partial \mathcal{F}(\hat{t})}{\partial \theta}),$$

where $\mathcal{F}(\cdot)$ is the free energy of RBM defined as

$$\mathcal{F}(t) = -\ln \sum_{h} e^{-E_{\mathrm{dRBM}}(t,h)}$$

Algorithm 1: Discriminative training for RBM

Input:
D: the training data set
GEN(a): N-best list for an utterance t in the reference
λ: learning rate
for *k=1:K* **do**
  for $a \in D$ **do**
    Positive:
    $\hat{t} = \mathtt{argmin}_{t' \in \mathrm{GEN}(a)} \mathrm{WER}(t')$
    Negative:
    $T^- = \{t' | 1 + Score(t') > Score(\hat{t})\}, t' \in \mathrm{GEN}(t)\}$
    for $t' \in T^-$ **do**
      $\theta \leftarrow \theta + \lambda \frac{\partial - \mathcal{F}(\hat{t})}{\partial \theta}$
      $\theta \leftarrow \theta - \lambda \frac{\partial - \mathcal{F}(t')}{\partial \theta}$

The training algorithm is described in Algorithm 1. For each acoustic input in training set, we select a set of hypotheses T^-, which are ranked higher than the oracle best hypothesis. Based on our analysis of the loss function, we boost the score of oracle best with its derivate of negative free energy and penalize hypothesis in T^- with their derivates of negative free energy. Note that, as compared with standard standard RBM training [15], which iterates over input space or samples of input space, our discriminative training needs only to iterate over the N-best list which grows linearly with the size of training data and N-best list.

2.3 Training with Prior Knowledge

The binary hidden layer of RBM allows for easily incorporating prior knowledge into the reranking model. We consider using two types of prior knowledge: named entity labels and pretrained latent layer from texts. Firstly, to improve the capability of recognizing content words, we capture prior of a special class of content words – named entities. As entity related prior also encodes information about word classes, it helps improving the generalization power of language models [16] as well.

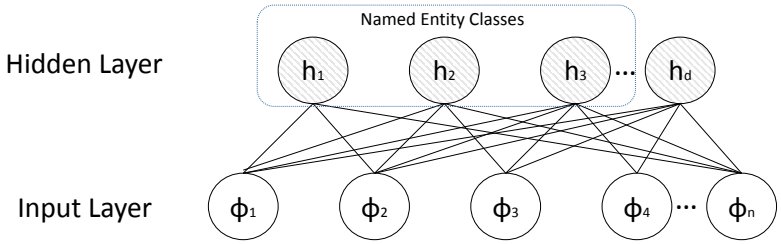

Fig. 1. Structure of RBM with Entity-related Prior

Specifically, we extract pairs of named entity words and their classes from texts using a named entity tagger, which annotates the text with 3 widely-adopted named entity classes, i.e., LOCATION, ORGANIZATION and PERSON. As show in Fig. 1, 3 variables in the hidden layer of RBM are used to represent named entity classes. For purpose of reducing ambiguity, we remove words belonging to multiple entity classes. We denote the list of entity-class pairs as $G = \{w, e\}$, where w is an index of the unigram feature in the input layer, and e an index of entity-class variable in the hidden layer. The objective function is then augmented with an entity-related regularizer,

$$L - \lambda \ln \prod_{w,e \in G} \prod (P(h_e = 1|\phi_w) - 1)^2,$$
$$P(h_e|\phi_w) = \sigma(c_e + W_{e,w}\phi_w).$$

As introduced in Wang et al. [12], $P(h_e|\phi_w)$ denotes the probability of a hidden variable h_e being activated by a given input feature ϕ_w.

To handle the data sparsity, we initiate connection matrix W of RBM with values pretrained using a large text corpora and the generative training. The pretraining captures the distributional semantics of input features [17].

2.4 Scoring ASR Hypotheses

To score a given hypothesis, we propose two scoring functions using our RBM-based reranker and its combination with SLP. First of all, the RBM-based reranking score S_{RBM} is defined as the logarithm of the unnormalized probability $\tilde{P}_{\text{dRBM}}(t)$ assigned by the RBM-based reranker solely,

$$
\begin{aligned}
S_{\mathrm{RBM}}(t) &= \ln \tilde{P}_{\mathrm{dRBM}}(t) \\
&= \underbrace{w_0 \ln(P(t|a))}_{\textbf{ASR posterior}} + \underbrace{\sum_i^n b_i \phi_i(t)}_{\textbf{linear part}} + \underbrace{\sum_j^d \ln(1 + e^{(c_i + W_i \phi(t))})}_{\textbf{hidden variable part}} .
\end{aligned}
$$

As shown above, the re-scoring function is composed of the original ASR posterior, a linear bias, and the hidden variable component. In addition, SLP and RBM are likely to have encoded information complementary to each other due to their different structures and training methods. Therefore, we propose a late fusion of the two methods, which combines their confidence scores in the testing phase. The combined reranking score is

$$S(t) = S_{\mathrm{RBM}}(t) + \alpha S_{\mathrm{SLP}}(t),$$

where $S_{\mathrm{SLP}}(t)$ is the single perceptron based confidence score weighted by α.

3 Related Work

DLM has been first introduced by Roark et al. in [6], where simple features like N-gram was shown able to effectively reduce WER. This previous work is using a Single Layer Perceptron (SLP) to modify the original posterior probabilities of the outputs of a baseline ASR system using a linear function,

$$\log P(t|a) + \sum_i w_i f_i(t),$$

where $\log P(t|a)$ is the log probability of a word sequence t given the acoustic signal a, and $\{f_i(\cdot)\}$ are the set of feature functions of an utterance weighted by $\{w_i\}$. Different types of features extracted from syntactic trees [7] and dependency trees [8] have also been used to enrich the feature set.

Apart from feature engineering and using linear combination of feature functions, inferring hidden variables from the observed input captures semantic information related to word classes and word senses. Our work is closely related with Koo et al. [4] who proposed a hidden-variable model to rerank syntactic parsing trees. For the tractability of their model, they put constrains on the connections between latent variables and visible variables (i.e., input layer) by splitting features into two sets. However, the way they divide features is specific to syntactic parsing, and thus is not applicable to our task. Our model differs from Koo et al. [4] in the sense that the connection is not constrained by their feature type, but instead relying on the structure of RBM to build connections between input and hidden layer.

RBM-based models [10, 11] have been explored for language modeling. Both approaches model the probability of a fixed length of word sequences, i.e., N-grams, and trained with a generative objective function. Our method differs from these methods in two major aspects: the training of proposed RBM reranker is discriminative, and it represents sentences of variable length as global feature vectors.

4 Experiment

4.1 Dataset

We evaluate our work on the latest release of TedLium Corpus [18] which is a set of audio and manually transcribed texts of Ted talks. As shown in Table 1. We split the training set of TedLium Version 2 into two parts: former Tedlium Training set Version 1 and the rest. The Version 1 part is a set of 774 Ted talks consisting of 56,800 utterances and more than 1.7 million words, while the remaining of the TedLium training set contains another 718 talks. The evaluation set of our experiment is the testing set of TedLium corpus, which is composed of 11 talks. Our text corpus is the ukWaC corpus [19], which is a collection of texts containing about 1.8 billion words.

	Utterances	Talks	Words
ASR AM Train	56.8K	774	1.7M
Reranking Train(speech)	36.2K	718	0.9M
Reranking Train(text)	24M	-	1.8B
Reranking Test	1.15K	11	29K

Table 1. Characteristics of the data sets used in experiments

4.2 Baseline

The baseline ASR system is based on KALDI[1] toolkit [20] including a DNN-based acoustic model. It uses a pre-trained language model[2], which is released as part of Sphinx project [21] and has achieved a perplexity of 158.3 on a corpus of Ted Talks. The acoustic model is trained using the training set of TedLium Version 1. The rest of training set of TedLium Version 2 is used for training reranking models. The baseline SLP reranker is trained by following the work of Lambert et al. [8] that randomly selects K pairs of hypotheses from the N-best lists. Specifically, we randomly select 100 pairs from the 100-best list. Learning process is ran for 10 iterations, as we cannot observe further WER reduction with more iterations.

[1] http://kaldi.sourceforge.net/
[2] http://cmusphinx.sourceforge.net/2013/01/a-new-english-language-model-release/

4.3 RBM Setup

We refer to the system integrating prior knowledge with dRBM as p-dRBM. Both dRBM and p-dRBM use 200 hidden units and are trained using the same data set and 100-best hypotheses as SLP reranker. Since our focus is not on feature engineering, and for simplicity of interpreting our experiment results, we use only unigram features (i.e., single words) in our experiments, which have also been shown as the most effective features by previous works [8]. For training p-dRBM, we first crawled down a set of text summaries of ted talks from Ted website. We then create a list of 20 words for each entity category by tagging the collected text summaries with Stanford named entity recognition tool [22] following description in section 2.3. A basic RBM is trained using ukWaC corpus and used for initiating the connection matrix W of p-dRBM. The late fusion of SLP and our proposed methods are denoted as SLP + dRBM and SLP + p-dRBM. We use $\lambda = 0.01$ and $\alpha = 1.0$ as weights of entity-related regularizer and SLP scores in late fusion.

4.4 Evaluation

First of all, we analyze the behavior of p-dRBM by computing the most-activated words by p-dRBM as shown in Table 2. As shown in previous section, the scoring function used by our method is a combination of ASR confidence score, a linear component (denoted as p-dRBM-L) and a hidden-variable component (denoted as p-dRBM-H). p-dRBM then takes as input one-hot vectors of words to compute their reranking scores. It shows that the linear component is mainly accounting for the function words, while the hidden component favors content words that are mostly nouns and adjectives. The final scoring function is a trade-off between function words and content words through a combination of the two components.

p-dRBM-H	**p-dRBM-L**	**p-dRBM**
integrated	of	integrated
demeanor	and	demeanor
disgust	the	disgust
tattoo	to	tattoo
formula	a	formula

Table 2. Most activated word by p-dRBM

To investigate on what is captured by RBM and potentially effective for improving the Word Error Rate (WER), we represent each hidden variable as a vector of words. These vectors represent how much a word is activated by a given hidden variable. Table 3 shows a selected set of hidden variables that can be seen as a set of topics. We found that RBM can capture meaningful topics by using only sentence-level co-occurrence. We then represent each word as a vector of hidden variables by taking the rows of the matrix $W \in R^{|V| \times d}$ of p-dRBM.

working	media	higher education	entertainment
security	news	cambridge	scene
services	forum	mary	story
office	business	professor	tv
home	press	william	songs
for	new	royal	moving

Table 3. Example topics learned by p-dRBM

We rank words based on their cosine similarity with the queries and select the top 5 words for four query words. As shown in Table 4, the top ranked words all seem very relevant to the query words. Since our RBM is trained with sentence-level concurrence, which is different from the window-based methods, the 'similarity' looks more like a topical relatedness rather than syntactical similarity. In general, we can conclude that the resultant RBM-based reranking model to some extent captures the distributional semantics related to the topics of words.

japan	film	bible	computer
india	story	greatest	software
italy	music	holy	database
asia	beautiful	truth	digital
germany	famous	gospel	user
china	classic	spirit	server

Table 4. Most-similar words for queries using p-dRBM word embeddings

As shown in Table 5, we evaluate WER of reranking systems. It shows that the proposed discriminatively trained RBM produces greater WER reduction than baseline SLP rerankers. Effectiveness of using prior knowledge is validated by further improving WER over dRBM. The greatest absolute WER reduction (1.3%) is achieved by the late fusion of SLP and p-dRBM, which confirms that our reranker captures information complementary to SLP.

	WER	**WER (TF-IDF$\geq$ 3)**
ASR 1-best	18.23	46.9
Oracle 1-best	11.42	36.1
SLP	17.76	46.3
dRBM	17.51	44.6
p-dRBM	17.36	**43.8**
SLP + dRBM	17.11	45.2
SLP + p-dRBM	**16.91**	44.2

Table 5. Performance of reranking model on TedLium corpus

Since the latent layer of p-dRBM incorporates prior knowledge related to content words (e.g., named entities), it is desirable that the proposed method can better recognize content words, which are more critical for downstream applications such as spoken language understanding. To evaluate the performance of our proposed methods on recognizing content words in a more general way, we words that have higher TF-IDF scores are more likely to be content words. We hence assign more weight to errors involving a set of keywords with high TF-IDF instead of treating all words equally. Specifically, the list of keywords are chosen based on TF-IDF scores (≥ 3.0) computed from whole TedLium corpus. We use the weighted-word-scoring implementation in NIST SCLITE tool [3] by aggressively assigning weight 1.0 to words on the list and 0.0 to the rest.

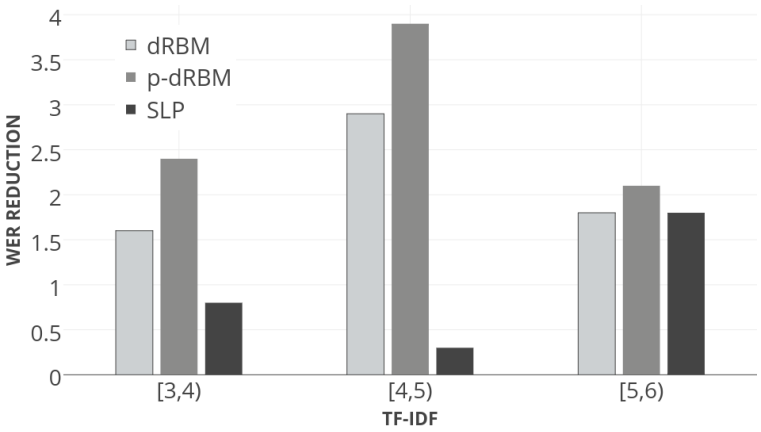

Fig. 2. WER reduction for words versus TF-IDF scores

Table 5 clearly shows that baseline reranking systems (SLP) fail to reduce much WER for selected keywords. In comparison, proposed RBM rerankers, especially p-dRBM, have reduced more errors on chosen keywords without sacrificing overall performance. We further break down the TF-IDF scores into 3 bins. Figure 2 shows the WER reduction by all three approaches. Thanks to hidden variables, our methods are capable of better capturing the discriminative information for most content words. In particular, p-dRBM is shown working significantly better than other methods on words with medium TF-IDF scores (<5), which is a result of injecting named entity words, e.g., washington (TF-IDF$= 3.87$), that mostly have TF-IDF between 3.0 and 5.0.

[3] http://www1.icsi.berkeley.edu/Speech/docs/sctk-1.2/sclite.htm

5 Conclusion

In this paper, we proposed an RBM-based language model that is discriminatively trained for reranking ASR hypotheses. In comparison with single perceptron based reranker, our proposed approach reduces more word errors. The success of fusing single perceptron and RBM-based reranker suggests that two models actually capture complementary information useful for selecting less erroneous ASR hypotheses. In addition, we found that introducing prior knowledge to RBM-based reranker results in a better recognition of content words. In the future, we would like to explore the use of lexical knowledge obtained from different resources, e.g., WordNet [23] or SenticNet [24], as additional prior knowledge for the proposed model.

References

1. Collins, M.: Ranking algorithms for named-entity extraction: Boosting and the voted perceptron. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. (2002) 489–496
2. Ma, Y., Cambria, E., Gao, S.: Label embedding for zero-shot fine-grained named entity typing. In: COLING, Osaka (2016) 171–180
3. Collins, M., Koo, T.: Discriminative reranking for natural language parsing. Comput. Linguist. **31** (2005) 25–70
4. Koo, T., Collins, M.: Hidden-variable models for discriminative reranking. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada (2005) 507–514
5. Li, Z., Khudanpur, S.: Large-scale discriminative n-gram language models for statistical machine translation. In: Proceedings of AMTA. (2009)
6. Roark, B., Saraclar, M., Collins, M., Johnson, M.: Discriminative language modeling with conditional random fields and the perceptron algorithm. In: Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume, Barcelona, Spain (2004) 47–54
7. Collins, M., Roark, B., Saraclar, M.: Discriminative syntactic language modeling for speech recognition. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. (2005) 507–514
8. Lambert, B., Raj, B., Singh, R.: Discriminatively trained dependency language modeling for conversational speech recognition. In: INTERSPEECH. (2013) 3414–3418
9. Smolensky, P.: Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. Information Processing in Dynamical Systems: Foundations of Harmony Theory (1986) 194–281
10. Niehues, J., Waibel, A.: Continuous space language models using restricted boltzmann machines. In: IWSLT. (2012) 164–170
11. Dahl, G.E., Adams, R.P., Larochelle, H.: Training restricted boltzmann machines on word observations. arXiv preprint arXiv:1202.5695 (2012)
12. Wang, L., Liu, K., Cao, Z., Zhao, J., de Melo, G.: Sentiment-aspect extraction based on restricted boltzmann machines. In: Proceedings of the 53rd Annual

Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, Association for Computational Linguistics (2015)
13. Fries, C.C. In: The Structure of English. Harcourt Brace, New York (1952)
14. Pernkopf, F., Wohlmayr, M., Tschiatschek, S.: Maximum margin bayesian network classifiers. IEEE Trans. Pattern Anal. Mach. Intell. **34** (2012) 521–532
15. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. Neural Computation **14** (2002) 1771–1800
16. Michael Levit, Sarangarajan Parthasarathy, S.C.: Word-phrase-entity language models: Getting more mileage out of n-grams. In: Proc. Interspeech, Singapore, ISCA - International Speech Communication Association (2014) 666–670
17. Salakhutdinov, R., Hinton, G.E.: Replicated softmax: an undirected topic model. In: NIPS. Volume 22. (2009) 1607–1614
18. Rousseau, A., Deléglise, P., Estève, Y.: Enhancing the ted-lium corpus with selected data for language modeling and more ted talks. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). (2014) 3935–3939
19. Ferraresi, A., Zanchetta, E., Baroni, M., Bernardini, S.: Introducing and evaluating ukwac, a very large web-derived corpus of english. In: Proceedings of WAC-4. (2008)
20. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, IEEE Signal Processing Society (2011)
21. Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P., Woelfel, J.: Sphinx-4: A flexible open source framework for speech recognition. Technical report, Mountain View, CA, USA (2004)
22. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics (2005) 363–370
23. Fellbaum, C.: WordNet: An Electronic Lexical Database (Language, Speech, and Communication). The MIT Press (1998)
24. Cambria, E., Poria, S., Bajpai, R., Schuller, B.: SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In: COLING. (2016) 2666–2677