Identifying Motifs in Folktales using Topic Models

Folgert Karsdorp FOLGERT.KARSDORP@MEERTENS.KNAW.NL
Meertens Institute, Joan Muyskenweg 25, 1096 CJ Amsterdam, The Netherlands

Antal van den Bosch A.VANDENBOSCH@LET.RU.NL
Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

Keywords: L-LDA, Topic Model, motifs, multi-labeled corpora, ranking, Okapi BM25

Abstract

With the undertake of various folktale digitalization initiatives, the need for computational aids to explore these collections is increasing. In this paper we compare Labeled LDA (L-LDA) to a simple retrieval model on the task of identifying motifs in folktales. We show that both methods are well able to successfully discriminate between relevant and irrelevant motifs. L-LDA represents motifs as distributions over words. In a second experiment we compare the quality of these distributions to those of a simple baseline that ranks words using a TF·IDF weighting scheme. We show that both models produce representations that match relatively well to a manually constructed motif classification system used in folktale research. Finally we show that unlike L-LDA, this simple baseline is capable of representing abstract motifs as generalizations over more specific motifs.

1. Introduction

Without the wondering question "What makes your ears so big?", the story of *Little Red Riding Hood* does not feel complete. Likewise, every telling of *Cinderella* should contain a part about the glass slipper and a cruel stepmother who makes the heroine's life miserable. In folktale research such more or less obligatory passages are called motifs. They "have a power to persist in tradition" (Thompson, 1946) and are part of our collective cultural heritage. Motifs play a key role in the classification of folktales into folktale types. For instance, in the authoritative folktale type catalog *The Types of International Folktales* by Aarne, Thompson and Uther (henceforth: ATU catalog) (Uther, 2004) every tale type is accompanied by a sequence of motifs which are the primary descriptive units of that tale type.

The goal of our work is to automatically identify motifs in folktales. This can be cast as a multi-label classification task in which we attempt to assign a set of motifs to unseen, unlabeled folktales. The set of potential labels that can be assigned to a folktale is large, but certain motifs will be more strongly tied to the particular folktale. We therefore conceptualize our task as a ranking problem.

As discussed in more detail by Karsdorp et al. (2012) and illustrated by Figure 1, the motifs in the Dutch Folktale Database follow a power-law like distribution. Recent research makes a strong case for the use of statistical topic models for multi-label datasets with long-tail label distributions as opposed to discriminative methods (Rubin et al., 2012). In this paper we compare the performance of the supervised topic model Labeled LDA (L-LDA) (Ramage et al., 2009) to a 'simple' retrieval model that uses Okapi BM25 as its ranking function. The first question we would like to answer is: How well do both systems perform on a ranking task where the goal is to allocate the highest ranks to the most relevant motifs?

Topic models such as LDA represent topics as distributions over words. Many studies are devoted to methods that aim to measure the quality and interpretability of these topics, which may not be trivial given the unsupervised nature of LDA. However, we are in a position in which we can use predefined labels, as the motifs used in this study are part of a

Appearing in *Proceedings of BENELEARN 2013*. Copyright 2013 by the author(s)/owner(s).

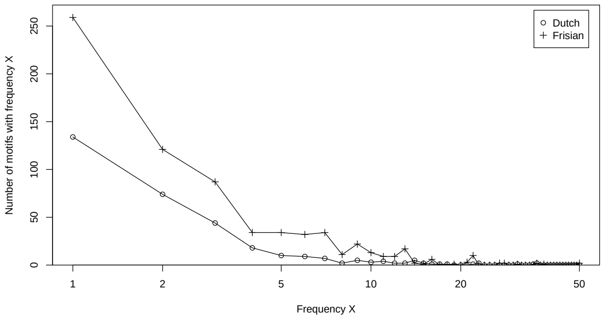

Figure 1. Frequency distribution of motifs on a log scale in Dutch and Frisian folktales in the Dutch Folktale Database.

(hierarchical) classification system, and we have information on which motifs occur in which folktale type. This information is available throughout our data, providing us with ground truth labels. We compare the motif representations discovered by L-LDA to those obtained using a simple baseline in which we compute what words are most strongly associated with each motif using a TF·IDF weighting scheme. We then verify by a quantitative evaluation (using several evaluation metrics from information retrieval) how well the motif representations discovered by both systems compare to a manually constructed motif classification system used in folktale research.

The automatic extraction of motifs is relevant for a number of reasons. Various new folktale digitization initiatives have been undertaken (Meder, 2010; Abello et al., 2012; La Barre & Tilley, 2012), which ask for ways to browse the collections at different facets, such as motifs. This would allow researchers to investigate, for example, how folktales have changed through time in terms of their motif material. It is only since the appearance of Brothers Grimm's version of *Little Red Riding Hood*, for example, that the girl and her grandmother are rescued from the wolf's belly. Extracting motifs from texts also allows researchers to find new relationships between folktales which could tell us more about their evolution.

The outline of the paper is as follows. We will start with providing an overview of related work in Section 2. We then continue with a description of the resources used in this study in Section 3. Sections 4 and 5 are devoted to the experimental setup followed by our results. The last section offers our conclusions and directions for future work.

2. Related work

Voigt et al. (1999) have shown that it is possible to automate motif identification in folklore text corpora by automatically grouping texts based on their content similarities. In their study, the presence of common motifs was derived from co-occurrences of keywords in the texts. For folklore researchers, however, the results are not easily interpretable because motifs are represented as principal components to which no label is assigned.

The literature on multi-label classification is very extensive and has been summarized elsewhere (e.g Tsoumakas & Katakis (2007)). Of special interest for our purposes is the recent work by Nguyen et al. (2013) who showed that Okapi BM25 acts as a competitive baseline in a folktale type classification experiment.

Our work is an application of the multi-label adaptation of Latent Dirichlet Allocation – Labeled LDA – as proposed by Ramage et al. (2009). Rubin et al. (2012) provide an extensive comparison of discriminative multi-label classifiers and three multi-labeled extensions of LDA. They make a strong case for the use of statistical topic models in the context of highly skewed datasets.

Our work differs from both aforementioned papers in three aspects. First, we apply the model to literary texts. It has been observed in many applications that literary texts behave differently from other genres in various ways which requires adaptations of the proposed models. Second, our multi-labeled dataset provides us with the unique possibility to evaluate the topic distributions against ground truth labels. Finally, we will propose a simple way to incorporate the hierarchical structure of our label set into the model.

3. Resources

3.1. TMI and ATU

The comprehensive *Motif-Index* (Thompson, 1955 1958) contains over 45,000 motifs. The motifs are hierarchically ordered in a tree structure. There are 23 alphabetic top-level categories ranging from mythological motifs to motifs concerning traits of character. Many motifs are bound to particular folktale types. Under (1) we list some examples:

(1)Q426 Wolf cut open and filled with stones as punishment;
F911.3 Animal swallows man (not fatally);
F823.2 Glass shoes.

The motifs from the TMI play a key role in the classification of tales into a certain type in the ATU catalog. Every folktale type contains a short summary of the plot. In this summary we find a sequence of motifs that together uniquely identify a folktale. An example of a story summary in the ATU catalog, of the folktale type *The Shepherd Boy*, is as follows.

> ATU 0515, "**The Shepherd Boy.** A foundling child who herds animals finds three objects (of glass) which he gives back to their owners. They promise to reward him [Q42]. With the help of the last owner, a giant, the boy fulfills three tasks. He acquires a castle in which a princess is confined. He rescues her and marries her [L161]."

This tale type contains two motifs, Q42 'Generosity rewarded' and L161 'Lowly hero marries princess'.

3.2. Dutch Folktale Database

The Dutch Folktale Database[1] is a collection of about 42,000 folktales (Meder, 2010). The collection contains folktales from various genres (e.g. fairytales, legends, urban legends, jokes) in a number of variants of Dutch and in Frisian. Every entry in the database contains metadata about the story, including language, collector, place and date of narration, keywords, names, and subgenre. The two largest components contain tales written in standard Dutch and Frisian. In this paper we restrict our experiments to these two components.

Folktales in the Dutch Folktale Database have been manually classified according to the folktale types in the ATU catalog, as far as a link could be established. This link between particular instances of tales and folktale types provides us with the set of motifs that can occur in a folktale type, and therefore in its instantiations. For each folktale in the Dutch Folktale Database that was classified according to the system in the ATU catalog, we assigned to it the set of motifs of its corresponding folktale type.

3.3. Datasets

We created two datasets: one for Dutch folktales and one for the Frisian tales. We only included tales that were classified according to the classification system of the ATU catalog. This resulted in 1,098 Dutch tales and 1,373 Frisian tales. Excluding punctuation, the average number of words per story is 468 for Dutch and 194 for Frisian. Both collections were tokenized using the Unicode tokenizer Ucto (Van Gompel et al., 2012).[2] We removed all diacritics and excluded words shorter than two characters and all numbers. As there are no off-the-shelf stemmers available for Frisian, we choose to not do any further preprocessing on the Dutch texts either and use the full tokens.

4. Models

4.1. Baselines

As a baseline for the Dutch and Frisian experiments we use a Big Document Model (see e.g. Nguyen et al. (2013)). For each motif observed in the collection we merge all documents in which that motif occurs into one big document. The ID number of the motif forms the class label of the new document. Given these big documents, we then compute the TF·IDF for all words. We use L2 to normalize the term vectors and smooth the IDF weights by adding one to the document frequencies. This provides us with a ranked list of how strongly a word is associated with a big document, i.e. a motif. We use these ranked lists as a baseline in the cluster evaluation in section 5.2. We will refer to this model as the Big Document Model (BDM).

As a baseline for the ranking experiment in section 5.1 we use a standard retrieval model with Okapi BM25 as our ranking function. BM25 has proven itself to be one of the most successful ranking functions in text-retrieval (Robertson & Zaragoza, 2009). We compute it as follows:

$$S(D,Q) = \textstyle\sum_{i=1}^{n} \mathrm{IDF}(q_i) \cdot \frac{f(q_i,D)\cdot(k_1+1)}{f(q_i,D)+k_1\cdot(1-b+b\cdot\frac{|D|}{\mathrm{avgdl}})} \quad (2)$$

where Q represents a query and $f(q_i, D)$ is the frequency of the i'th term in q in document D. Avgdl is the average document length. The parameters b and k_1 are set to 0.75 and 1.2, respectively. We compute the IDF weight using:

$$IDF(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} \quad (3)$$

where N is the number of documents in the corpus and $n(q_i)$ the number of documents that contain q_i. This formulation of IDF can result in negative scores when terms appear in more than fifty percent of the documents. We therefore give the summand in (2) a floor of zero, to filter common terms.

Queries are represented by the complete contents of a test folktale. We issue these queries on the constructed

[1] http://www.verhalenbank.nl

[2] http://ilk.uvt.nl/ucto/

big documents, resulting in a ranking of motifs for that particular folktale.

4.2. Labeled LDA

Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a popular method for extracting topics from texts. LDA is a generative probabilistic model that models documents as distributions over topics. Topics are represented by distributions over words. The model assumes that each word in a document is generated from a single topic.

Ramage et al. (2009) extend the basic framework of LDA by introducing a supervised variant in which the latent topics in LDA correspond directly to the labels assigned to a particular document. Given a corpus of multi-labeled documents the model can estimate the most likely words per label as well as the distribution of labels per document. The primary goal of Ramage et al. (2009) is to show what qualitative advantages L-LDA has over 'traditional' discriminative multi-labeled classifiers such as SVMs. Their results suggest that L-LDA might be advantageous in the context of highly skewed multi-labeled datasets, such as our corpus of folktales (see Rubin et al. (2012) for a more extensive comparison between multi-labeled supervised versions of LDA and SVM classifiers).

In the generative model of L-LDA labels are assumed to be generated from a binomial distribution. As Rubin et al. (2012) point out, in practice L-LDA just assumes the labels to be observed without a prior generative process. For educative purposes they propose a new model – Flat LDA – that does away with this assumption. Our implementation of the model is based on Flat LDA. However, we will still call the model Labeled LDA.

Unlike in unsupervised LDA, we are confident about the labels assigned to a document. To reflect this knowledge, and in order to reduce the variance of the topic distributions, we assign to the labels a relatively high prior ($\alpha = 50$). Because of the relatively small vocabulary size of our corpus, we use a relatively low term smoothing prior ($\beta = 0.001$) to assign the probability mass to only a few words per topic. Both α and β are symmetric priors.

5. Experimental results

5.1. Ranking experiment

In this section we will investigate to what extent we can use L-LDA as a multi-label classifier for the extraction of motifs. We cast the assignment of a set of labels to a document as a ranking task in which the goal is to allocate the highest ranks to the most relevant motifs. We rank the motifs according to their posterior probability in a document. We compare the performance of L-LDA to the retrieval model as described in section 4.1.

Table 1. Evaluation of motif retrieval for BM25 and L-LDA on Dutch and Frisian folktales.

	Model	AP	One Error	Is Error	Margin
Dutch	BM25	0.78	0.26	0.27	10.69
	L-LDA	0.72	0.30	0.39	26.48
Frisian	BM25	0.88	0.15	0.15	4.46
	L-LDA	0.88	0.16	0.16	7.0

We performed 10-fold cross-validation on both datasets, dividing the folktales at random into 10 parts of approximately equal size. As shown by Karsdorp et al. (2012), there are quite many pairs of motifs that co-occur exclusively, that is, they never appear without the other. For these motifs, both models have no way of knowing which words are relevant to which motif as – in information theoretic terms – their mutual information is maximal. We therefore choose to exclude all these informationally indistinguishable motifs from our experiments. Although this results in a rather drastic filtering of motif types, the final number of motif types is still sufficiently high (Frisian: 155, Dutch: 179) and still about eight times higher than in the experiments by Ramage et al. (2009).

In the ideal case, the top of the ranked list contains the motifs of a folktale. The extent to which this is the case reflects how well relevant motifs are found by the systems. We evaluate the ranked lists by means of four evaluation metrics (for reasons of comparability we follow Rubin et al. (2012) in our choice for these evaluation metrics):

Average Precision – Are most or all of the target motifs high up in the ranking?

One Error – For what fraction of documents is the highest-ranked motif incorrect?

Is Error – What fraction of rankings is not perfect?

Margin – What is the absolute difference between the highest ranked irrelevant motif and the lowest ranked relevant motif, averaged across folktales?

The results presented in Table 1 show quite similar results for both L-LDA and BM25. Surprisingly, the relatively standard retrieval model performs best on all evaluation metrics and on both datasets. In the case of the Frisian folktales the retrieval system is able to emphasize the highest ranks with high precision and a low irrelevance margin. L-LDA produces similar scores but has a slightly higher margin score. Both systems perform better on the Frisian tales than on the Dutch tales. Part of the explanation for this lies in the ratio between motifs and tales in the Dutch collection: there are relatively few folktales with many possible motifs, while the Frisian data has a higher average number of motifs per tale. BM25 shows less sensitivity to this ratio than L-LDA and outperforms L-LDA clearly. In the next section we perform a qualitative analysis to explore why this is the case when we evaluate the motif representations discovered by both models.

5.2. Motif visualization and evaluation

We compare the word distributions discovered by L-LDA to those found by the Big Document Model in which we compute the TF·IDF score for all words in each document. Table 2 shows the top words associated with four motifs for L-LDA and the BDM extracted from Dutch texts (the words are given in their English translation). Many words are discovered by both systems; especially the first few words are found by both methods. However, in some cases L-LDA misses some words characteristic of the given motif. Take motif N211.1.3, ‘Lost ring found in fish.’ L-LDA ranks the words *fish* and *ring* considerably lower than the BDM.[3]

Standard evaluation of topic identification by LDA is done on the basis of either extrinsic methods (such as retrieval tasks) or intrinsic methods, where the goal is to estimate the probability of test documents or to compute the coherence of topics (see Mimno et al. (2011) and the references cited therein). A rather unique property of the labels under investigation in this study is that they are part of a hierarchical tree structure. A motif such as ‘Transformation: pumpkin to carriage’ (D451.3.3) belongs to the more abstract category of ‘Transformation: object to object’ (D450–D499) which in turn is a child motif of the broader parent motif ‘Transformation’ (D0–D699), which in turn is placed under the top-level node ‘Magic’ (D), one out of the 23 top nodes.

We perform a hierarchical cluster analysis on the basis of the motifs discovered by L-LDA and evaluate the clusters against the top 23 categories in the hierarchical tree structure of Thompson’s *Motif Index*. We choose Ward’s method as our linkage method and compute the similarity between motifs using the cosine similarity metric.

Table 3. Clustering results of Dutch and Frisian motif representations.

	Model	homogeneity	completeness	V-measure
Dutch	BDM	0.365	0.330	0.347
	L-LDA	0.344	0.281	0.310
Frisian	BDM	0.354	0.299	0.324
	L-LDA	0.358	0.270	0.308

We evaluate the cluster solution on the basis of three measures (Rosenberg & Hirschberg, 2009):

Homogeneity – Does the cluster solution result in clusters that *only* contain members of the same class?

Completeness – Does the cluster solution result in clusters to which *all* members of the same class have been assigned?

V-measure – An entropy-based measure that expresses the harmonic mean of homogeneity and completeness.

The results in Table 3 show that the quality of the cluster solutions of the two models is quite similar. The solution obtained from the BDM corresponds slightly better to the top-level categorization in the *Motif Index* than the one from L-LDA.

5.3. Exploiting the hierarchical structure of the Motif Index

In the model described above the set of possible motifs was restricted to those motifs that are present in the training data. In the following we describe an extension of the model in which we exploit the relations between motifs in the hierarchical tree of Thompson’s *Motif Index*, which lists many motifs not present in the ATU catalog. Yet, because of the hierarchical nature of the index, many ancestral motifs are implicitly observed. The question we would like to explore is: What can we learn about the representation of these more abstract motifs by exploiting the hierarchical structure of the index?

[3] It is not necessarily a ‘ring’ that is found in the fish. There are many variations on this folktale type and often ‘teeth’ or a ‘denture’ is found in the fish’ belly, which is why BDM ranks these words so high.

Table 2. The top words within four motifs learned by L-LDA and BDM.

TD·IDF	L-LDA
Q426: Wolf cut open and filled with stones as punishment.	
wolf, Little Red Riding Hood, grandmother, her, children, little kids, your, Oud-Bovetje, big, granny, belly, goat, mother, angry	*wolf, children, mother, door, said, open, her, little kids, so, entire, still, belly, surely, Oud-Bovetje, went*
N211.1: Lost ring found in fish.	
Stavoren, teeth, cod, her, denture, ring, sea, wheat, ships, fish, shipper, harbor, she, the Heerhugowaard	*the, her, and, she, the, of, in, was, a, lady, Stavoren, she, ring, sea, denture*
K343.2.1: The stingy parson and the slaughtered pig.	
clerk, pastor, pig, stolen, will, slaughter, farmers, tonight, everyone, belief, sexton's house, fattened, excellent, slaughter time, pig meat, insignificant	*clerk, pastor, pig, will, said, asked, everyone, stolen, mine, yes, so, must, against*
J2321.1: Parson made to believe that he will bear a calf.	
student, pastor, little bottle, cork, uroscopy, monkey, John, clerk, pregnant, rubber band, quack, butt, give birth, your, spins	*the (de), a, pastor, John, student, the (het), to be, must, water, says, comes, to (te), to (om), and, surely*

Table 4. Clustering results of Dutch and Frisian motif representations (including ancestor motifs).

	Model	homogeneity	completeness	V-measure
Dutch	BDM	0.339	0.315	0.327
	L-LDA	0.159	0.177	0.168
Frisian	BDM	0.414	0.377	0.394
	L-LDA	0.197	0.199	0.198

Figure 2 shows the tale type ATU 333 *Little Red Riding Hood* as a layered sequence of motifs. The gray nodes are observed in the ATU catalog under index ATU 333. The observed motifs inherit certain information from its ancestors. Although we have no direct information about the unshaded motifs in the graph, it should be possible to infer some information about their features. The motifs F911.3 and F913, for example, share the concept of "extraordinary swallowing" and have some idiosyncratic aspects themselves. If we assume that a motif such as F911.3 is a mixture of features from its parents and of its own, we might be able to learn about the features of the unobserved more abstract motifs.

Each folktale is labeled with the motifs that are listed by its corresponding tale type in the ATU catalog. We expand this motif set by incorporating all ancestral motifs in Thompson's *Motif Index.* We only take into account non-terminal nodes with at least two children. The top-level categories in the index miss an overarching root node, which we add to the tree. Similar as before, we exclude all motifs from the experiment that exhibit maximal mutual information towards each other. This results in 410 possible motifs in the Dutch dataset and 293 motifs in the Frisian dataset.

Table 4 shows the evaluation of the cluster solutions. Interestingly, whereas in the previous evaluation L-LDA and BDM gave similar results, here L-LDA seems to suffer considerably from the addition of ancestral nodes to the observed motifs. The cluster solution obtained from BDM outperforms L-LDA by a substantial margin on all evaluation measures. To obtain a better intuition about why BDM performs better than L-LDA in matching its motif distributions to the hierarchy in Thompson's *Motif Index*, we show part of the hierarchical tree in Figure 3. We display for each motif the top words discovered by the two models. The words discovered by BDM are listed in the left column. The right column displays those of L-LDA.

Various interesting observations can be made on the motif representations in the tree. First, intuitively both L-LDA and BDM are able to discover good quality motifs at the leaves of the tree. Take motif J1780 'Things thought to be devils, ghosts etc.' where L-LDA is able to find some either directly or indirectly related words such as *child molester, butchery* and *world war.* BDM provides a good motif representation for J1150: 'Cleverness connected with the giving of evidence' with words such as *fish pot, fox trap* and *money.* All three items function as important pieces of evidence in the court of law in variants of ATU 1381 'The Talkative Wife and the Discovered Treasure'.

Inspecting the tree provides us with two hypotheses about why L-LDA performs much worse on the cluster evaluation than BDM. First, several motifs contain

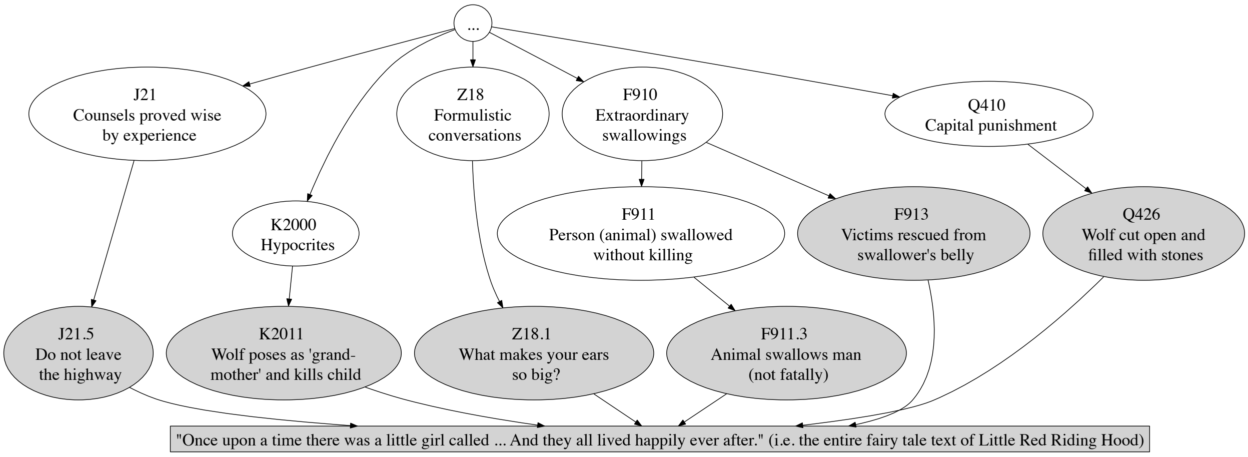
...
J21
Counsels proved wise by experience
Z18
Formulistic conversations
F910
Extraordinary swallowings
Q410
Capital punishment
K2000
Hypocrites
F911
Person (animal) swallowed without killing
F913
Victims rescued from swallower's belly
Q426
Wolf cut open and filled with stones
J21.5
Do not leave the highway
K2011
Wolf poses as 'grand-mother' and kills child
Z18.1
What makes your ears so big?
F911.3
Animal swallows man (not fatally)
"Once upon a time there was a little girl called ... And they all lived happily ever after." (i.e. the entire fairy tale text of Little Red Riding Hood)

Figure 2. Motif sequence in ATU 333 *Little Red Riding Hood* (gray nodes represent observed motifs), expanded with the ancestor motifs in Thompson's *Motif Index*.

many stop words, although we filtered all words that appear in more than fifty percent of all documents. These content-free words provide little to no clue to discriminate between motif categories, but in L-LDA they play a rather large role in contrast to BDM. The second reason for the superiority of BDM over L-LDA appears to be that BDM incorporates the knowledge from lower-level motifs into the more abstract motifs. A clear example of this is the top-level motif 'J: The wise and the foolish'. Almost all top words are examples of characters in stories that are either wise or foolish. We expanded the original motif set of a document with ancestral motifs. The result of this design choice is that the hierarchical relations between motifs are only implicitly present. Because L-LDA assigns each word in a document to a single motif, motifs that occur in only a few documents will attract more lexically specific words than their ancestors that appear in more documents. This 'restriction' does not apply to BDM, where the same word may be assigned to both lower-level and higher-level motifs. In sum, L-LDA is capable of finding good representations of motifs, but they seem unrelated and the knowledge from higher-level motifs is not inherited by their children.

6. Conclusion

In this paper we applied Labeled LDA to the domain of folktales. We have shown that L-LDA functions as a competitive method to identify motifs in folktales. However, it lags behind on a relatively simple retrieval model that uses Okapi BM25 as its ranking function.

We evaluated the quality of the motifs found by L-LDA and BDM against the most important motif classification system in folktale research. The results showed that both L-LDA and BDM are well capable of discovering high-quality motifs for the lowest-level motifs. However, the motif representation discovered by L-LDA for higher-level motifs are of low quality. In contrast, BDM is able to exploit the hierarchical relations between motifs. The more abstract motifs are in fact generalizations over lower-level ones.

One of the most interesting properties of LDA is that it assigns each word in a document to a single topic. As shown by Ramage et al. (2009), these word-by-word topic assignments could allow us to detect which parts of a text correspond to the tags assigned at the document level. Likewise, we could use this information to localize the specific places at which motifs occur in folktales. Future research should therefore be directed at improving the quality of motif representations as discovered by L-LDA or, in competition with L-LDA, the development of a system that incorporates the motif representations found by BDM, by finding those parts of a text that support a detected motif best.

Acknowledgments

The authors would like to thank the three anonymous reviewers, Theo Meder and Peter van Kranenburg for their helpful comments. The work on which this paper is based has been supported by the Computational Humanities Program of the Royal Netherlands Academy of Arts and Sciences, as part of the Tunes & Tales project.

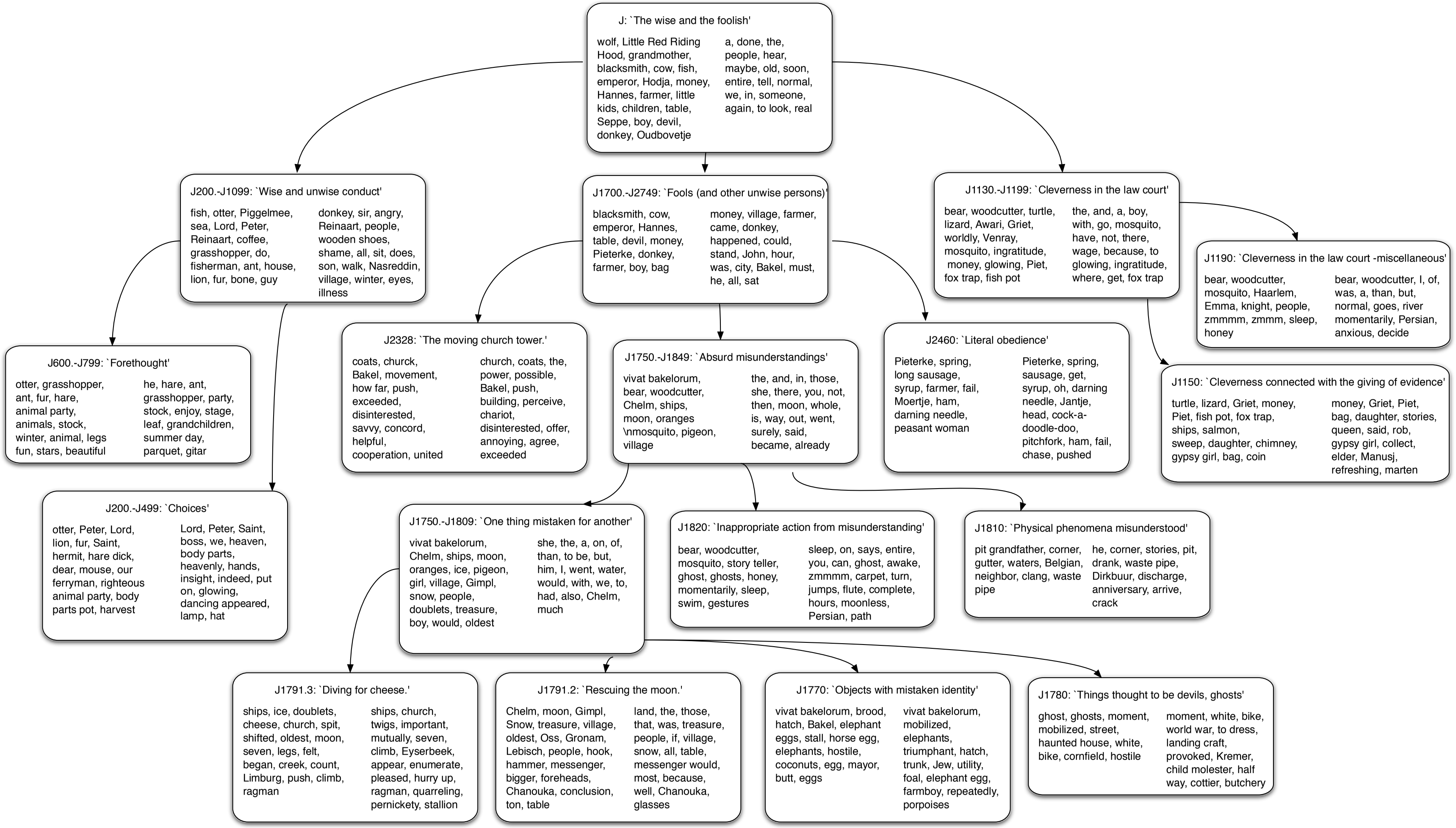
J: `The wise and the foolish'
wolf, Little Red Riding Hood, grandmother, blacksmith, cow, fish, emperor, Hodja, money, Hannes, farmer, little kids, children, table, Seppe, boy, devil, donkey, Oudbovetje
a, done, the, people, hear, maybe, old, soon, entire, tell, normal, we, in, someone, again, to look, real
J200.-J1099: `Wise and unwise conduct'
fish, otter, Piggelmee, sea, Lord, Peter, Reinaart, coffee, grasshopper, do, fisherman, ant, house, lion, fur, bone, guy
donkey, sir, angry, Reinaart, people, wooden shoes, shame, all, sit, does, son, walk, Nasreddin, village, winter, eyes, illness
J1700.-J2749: `Fools (and other unwise persons)'
blacksmith, cow, emperor, Hannes, table, devil, money, Pieterke, donkey, farmer, boy, bag
money, village, farmer, came, donkey, happened, could, stand, John, hour, was, city, Bakel, must, he, all, sat
J1130.-J1199: `Cleverness in the law court'
bear, woodcutter, turtle, lizard, Awari, Griet, worldly, Venray, mosquito, ingratitude, money, glowing, Piet, fox trap, fish pot
the, and, a, boy, with, go, mosquito, have, not, there, wage, because, to glowing, ingratitude, where, get, fox trap
J1190: `Cleverness in the law court -miscellaneous'
bear, woodcutter, mosquito, Haarlem, Emma, knight, people, zmmmm, zmmm, sleep, honey
bear, woodcutter, I, of, was, a, than, but, normal, goes, river momentarily, Persian, anxious, decide
J1150: `Cleverness connected with the giving of evidence'
turtle, lizard, Griet, money, Piet, fish pot, fox trap, ships, salmon, sweep, daughter, chimney, gypsy girl, bag, coin
money, Griet, Piet, bag, daughter, stories, queen, said, rob, gypsy girl, collect, elder, Manusj, refreshing, marten
J600.-J799: `Forethought'
otter, grasshopper, ant, fur, hare, animal party, animals, stock, winter, animal, legs fun, stars, beautiful
he, hare, ant, grasshopper, party, stock, enjoy, stage, leaf, grandchildren, summer day, parquet, gitar
J200.-J499: `Choices'
otter, Peter, Lord, lion, fur, Saint, hermit, hare dick, dear, mouse, our ferryman, righteous animal party, body parts pot, harvest
Lord, Peter, Saint, boss, we, heaven, body parts, heavenly, hands, insight, indeed, put on, glowing, dancing appeared, lamp, hat
J2328: `The moving church tower.'
coats, churck, Bakel, movement, how far, push, exceeded, disinterested, savvy, concord, helpful, cooperation, united
church, coats, the, power, possible, Bakel, push, building, perceive, chariot, disinterested, offer, annoying, agree, exceeded
J1750.-J1849: `Absurd misunderstandings'
vivat bakelorum, bear, woodcutter, Chelm, ships, moon, oranges \nmosquito, pigeon, village
the, and, in, those, she, there, you, not, then, moon, whole, is, way, out, went, surely, said, became, already
J2460: `Literal obedience'
Pieterke, spring, long sausage, syrup, farmer, fail, Moertje, ham, darning needle, peasant woman
Pieterke, spring, sausage, get, syrup, oh, darning needle, Jantje, head, cock-a-doodle-doo, pitchfork, ham, fail, chase, pushed
J1750.-J1809: `One thing mistaken for another'
vivat bakelorum, Chelm, ships, moon, oranges, ice, pigeon, girl, village, Gimpl, snow, people, doublets, treasure, boy, would, oldest
she, the, a, on, of, than, to be, but, him, I, went, water, would, with, we, to, had, also, Chelm, much
J1820: `Inappropriate action from misunderstanding'
bear, woodcutter, mosquito, story teller, ghost, ghosts, honey, momentarily, sleep, swim, gestures
sleep, on, says, entire, you, can, ghost, awake, zmmmm, carpet, turn, jumps, flute, complete, hours, moonless, Persian, path
J1810: `Physical phenomena misunderstood'
pit grandfather, corner, gutter, waters, Belgian, neighbor, clang, waste pipe
he, corner, stories, pit, drank, waste pipe, Dirkbuur, discharge, anniversary, arrive, crack
J1791.3: `Diving for cheese.'
ships, ice, doublets, cheese, church, spit, shifted, oldest, moon, seven, legs, felt, began, creek, count, Limburg, push, climb, ragman
ships, church, twigs, important, mutually, seven, climb, Eyserbeek, appear, enumerate, pleased, hurry up, ragman, quarreling, pernickety, stallion
J1791.2: `Rescuing the moon.'
Chelm, moon, Gimpl, Snow, treasure, village, oldest, Oss, Gronam, Lebisch, people, hook, hammer, messenger, bigger, foreheads, Chanouka, conclusion, ton, table
land, the, those, that, was, treasure, people, if, village, snow, all, table, messenger would, most, because, well, Chanouka, glasses
J1770: `Objects with mistaken identity'
vivat bakelorum, brood, hatch, Bakel, elephant eggs, stall, horse egg, elephants, hostile, coconuts, egg, mayor, butt, eggs
vivat bakelorum, mobilized, elephants, triumphant, hatch, trunk, Jew, utility, foal, elephant egg, farmboy, repeatedly, porpoises
J1780: `Things thought to be devils, ghosts'
ghost, ghosts, moment, mobilized, street, haunted house, white, bike, cornfield, hostile
moment, white, bike, world war, to dress, landing craft, provoked, Kremer, child molester, half way, cottier, butchery

Figure 3. Motif representations for part of Thompson's *Motif Index* found by BDM (left column) and L-LDA (right column).

References

Abello, J., Broadwell, P., & Tangherlini, T. R. (2012). Computational folkloristics. *Commun. ACM*, *55*, 60–70.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, *3*, 993–1022.

Karsdorp, F., Van Kranenburg, P., Meder, T., & Van den Bosch, A. (2012). In search of an appropriate abstraction level for motif annotations. *Proceedings of the 2012 Computational Models of Narrative Workshop* (pp. 22–26). Istanbul, Turkey.

La Barre, K. A., & Tilley, C. L. (2012). The elusive tale: leveraging the study of information seeking and knowledge organization to improve access to and discovery of folktales. *Journal of the American Society for Information Science and Technology*, *63*, 687–701.

Meder, T. (2010). From a dutch folktale database towards an international folktale database. *Fabula*, *51*, 6–22.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Edinburgh, Scotland.

Nguyen, D., Trieschnigg, D., & Theune, M. (2013). Folktale classification using learning to rank. *Advances in Information Retrieval, 35th European Conference on IR Research, ECIR 2013* (pp. 195–206). Moscow, Russia.

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 248–256). Singapore.

Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends in Information Retrieval*, *3*.

Rosenberg, A., & Hirschberg, J. (2009). V-measure: A conditional entropy-based external cluster evaluation measure. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 410–420). Prague.

Rubin, T. N., Chambers, A., Smyth, P., & Steyvers, M. (2012). Statistical topic models for multi-labeled document classification. *Mach Learn*, *88*, 157–208.

Thompson, S. (1946). *The folktale.* New York: Dryden Press.

Thompson, S. (1955–1958). *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jestbooks, and local legends.* Indiana University Press.

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, *3*, 1–13.

Uther, H.-J. (2004). *The types of international folktales: a classification and bibliography based on the system of antti aarne and stith thompson*, vol. 1–3 of *FF Communications.* Helsinki: Academia Scientarium Fennica.

Van Gompel, M., Van der Sloot, K., & Van den Bosch, A. (2012). *Ucto: Unicode tokeniser.* Radboud University Nijmegen / Tilburg University. Ilk technical report edition.

Voigt, V., Preminger, M., Ládi, L., & Darány, S. (1999). Automated motif identification in folklore text. *Folklore. An Electronic Journal of Folklore*, *12*, 126–141.