From Speaker Identification to Affective Analysis: A Multi-Step System for Analyzing Children's Stories

Elias Iosif[*] **and Taniya Mishra**[†]
[*] School of ECE, Technical University of Crete, Chania 73100, Greece
[†] AT&T Labs, 33 Thomas Street, New York, NY 10007, USA
iosife@telecom.tuc.gr, taniya@research.att.com

Abstract

We propose a multi-step system for the analysis of children's stories that is intended to be part of a larger text-to-speech-based storytelling system. A hybrid approach is adopted, where pattern-based and statistical methods are used along with utilization of external knowledge sources. This system performs the following story analysis tasks: identification of characters in each story; attribution of quotes to specific story characters; identification of character age, gender and other salient personality attributes; and finally, affective analysis of the quoted material. The different types of analyses were evaluated using several datasets. For the quote attribution, as well as for the gender and age estimation, substantial improvement over baseline was realized, whereas results for personality attribute estimation and valence estimation are more modest.

1 Introduction

Children love listening to stories. Listening to stories — read or narrated — has been shown to be positively correlated with children's linguistic and intellectual development (Natsiopoulou et al., 2006). Shared story reading with parents or teachers helps children to learn about vocabulary, syntax and phonology, and to develop narrative comprehension and awareness of the concepts of print, all of which are linked to developing reading and writing skills (National Early Literacy Panel 2008). While acknowledging that the parental role in storytelling is irreplaceable, we consider text-to-speech (TTS) enabled storytelling systems (Rusko et al., 2013; Zhang et al., 2003; Theune et al., 2006) to be aligned with the class of child-oriented applications that aim to aid learning.

For a TTS-based digital storytelling system to successfully create an experience as engaging as human storytelling, the underlying speech synthesis system has to narrate the story in a "storytelling speech style" (Theune et al., 2006), generate dialogs uttered by different characters using synthetic voices appropriate for each character's gender, age and personality (Greene et al., 2012), and express quotes demonstrating emotions such as sadness, fear, happiness, anger and surprise (Alm, 2008) with realistic expression (Murray and Arnott, 2008). However, before any of the aforementioned requirements — all related to speech generation — can be met, the text of the story has to be analyzed to identify which portions of the text should be rendered by the narrator and which by each of the characters in the story, who are the different characters in the story, what is each character's gender, age, or other salient personality attributes that may influence the voice assigned to that character, and what is the expressed affect in each of the character quotes.

Each of these text analysis tasks has been approached in past work (as described in our Related Works section). However, there appears to be no single story analysis system that performs all four of these tasks, which can be pipelined with one of the many currently available text-to-speech systems to build a TTS-based storyteller system. Without such a story analysis system, it will not be possible to develop an engaging and lively digital storyteller system, despite the prevalence of several mature TTS systems.

Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL) @ EACL 2014, pages 40–49,
Gothenburg, Sweden, April 27, 2014. ©2014 Association for Computational Linguistics

In this paper, we present a multi-step text analysis system for analyzing children's stories that performs all four analysis tasks: (i) Character Identification, i.e., identifying the different characters in the story, (ii) Quote Attribution, i.e., identifying which portions of the text should be rendered by the narrator versus by particular characters in the story, (iii) Character Attribute Identification, i.e., identifying each character's gender, age, or salient personality attributes that may influence the voice that the speech synthesis system assigns to each character, and (iv) Affective Analysis, i.e., estimating the affect of the character quotes.

This story analysis system was developed to be part of a larger TTS-based storyteller system aimed at children. As a result, the data used for developing the computational models or rules in each step of our system were obtained from children's stories. A majority of children's stories are short. They often contain multiple characters, each with different personalities, genders, age, ethnicities, etc., with some characters even being anthropomorphic, e.g., the singing candlestick or the talking teapot. In addition, there are several prototypical templates characterizing the main characters in the story (Rusko et al., 2013). However, character development is limited in these stories due to the shorter length of text. Overall, children's stories can be regarded as a parsimonious yet fertile framework for developing computational models for literature analysis in general.

2 Related Work

Elson and McKeown (2010) used rule-based and statistical learning approaches to identify candidate characters and attribute each quote to the most likely speaker. Two broad approaches for the identification of story characters were followed: (i) named entity recognition, and (ii) identification of character nominals, e.g., "her grandma", using syntactic patterns. A long list of heuristics for character identification is proposed in (Mamede and Chaleira, 2004). He et al. (2013) use a supervised machine learning approach to address the same problem, though many of their preliminary steps and input features are similar to those used in (Elson and McKeown, 2010). Our character identification and quote attribution is based on syntactic and heuristic rules that is motivated by each of these works.

There are two interesting sub-problems related to quote attribution. First is the problem of identifying anaphoric speakers, i.e., in the utterance *"Hello", he said*, which character is referred to by the pronoun *he*? This problem is addressed in (Elson and McKeown, 2010) and (He et al., 2013) but not in (Mamede and Chaleira, 2004). The second problem is resolving utterance chains with implicit speakers. Elson and McKeown (2010) describe and address two basic types of utterance chains: (i) one-character chains, and (ii) intertwined chains. In these chains of utterances, the speaker is not explicitly mentioned because the author relies on the shared understanding with the reader that adjacent pieces of quoted speech are not independent (Zhang et al., 2003; Elson and McKeown, 2010). They are either a continuation of the same character's speech (one-character chains) or a dialogue between the two characters (intertwined chains). In (Zhang et al., 2003), the quote-identification module detects whether a piece of quoted speech is a new quote (NEW), spoken by a speaker different from the previous speaker, or a continuation quote (CONT) spoken by the same speaker as that of the previous quote. He et al. (2013) also identified similar chains of utterances and addressed their attribution to characters using a model-based approach. In this work, we address both subproblems, namely, anaphoric speaker and implicit speaker identification.

Cabral et al. (2006) have shown that assigning an appropriate voice for a character in a digital storyteller system is significant for understanding a story, perceiving affective content, perceiving the voice as credible, and overall listener satisfaction. Greene et al. (2012) have shown that the appropriateness of the voice assigned to a synthetic character is strongly related to knowing the gender, age and other salient personality attributes of the character. Given this, we have developed rule-based, machine-learning-based and resource-based approaches for estimation of character gender, age and salient personality attributes. In contrast, the majority of past works on the analysis of children stories for TTS-based storytelling is limited to the attribution of quotes to speakers, though studies that focused on anaphoric speaker identification have also approached character gender estimation such as (Elson and McKeown, 2010) and (He et al., 2013). The utilization of available resources containing associations between person names and gender was followed in (Elson and

McKeown, 2010). In (He et al., 2013), associations between characters and their gender were performed using anaphora rules (Mitkov, 2002).

There is of course a significant body of work from other research areas that are related to the estimation of character attributes, similar to what we have attempted in our work. Several shallow linguistic features were proposed in (Schler et al., 2006) for gender identification, applied to the identification of users in social media. Several socio-linguistic features were proposed in (Rao et al., 2010) for estimating the age and gender of Twitter users. The identification of personality attributes from text is often motivated by psychological models. In (Celli, 2012), a list of linguistic features were used for the creation of character models in terms of the the Big Five personality dimensions (Norman, 1963).

Analysis of text to estimate affect or sentiment is a relatively recent research topic that has attracted great interest, as reflected by a series of shared evaluation tasks, e.g., analysis of news headlines (Strapparava and Mihalcea, 2007) and tweets (Nakov et al., 2013). Relevant applications deal with numerous domains such as blogs (Balog et al., 2006), news stories (Lloyd et al., 2005), and product reviews (Hu and Liu, 2004). In (Turney and Littman, 2002), the affective ratings of unknown words were predicted using the affective ratings for a small set of words (seeds) and the semantic relatedness between the unknown and the seed words. An example of sentence-level analysis was proposed in (Malandrakis et al., 2013). In (Alm et al., 2005) and (Alm, 2008), linguistic features were used for affect analysis in fairy tales. In our work, we employ a feature set similar to that in (Alm et al., 2005). We deal with the prediction of three basic affective labels which are adequate for the intended application (i.e., storytelling system), while in (Alm, 2008) more fine-grained predictions are considered.

The integration of various types of analysis constitutes the distinguishing character of our work.

3 Overview of System Architecture

The system consists of several sub-systems that are linked in a pipeline. The input to the system is simply the text of a story with no additional annotation. The story analysis is performed sequentially, with each sub-system extracting specific information needed to perform the four analysis tasks laid out in this paper.

3.1 Linguistic Preprocessing

The first step is linguistic pre-processing of the stories. This includes (i) tokenization, (ii) sentence splitting and identification of paragraph boundaries, (iii) part-of-speech (POS) tagging, (iv) lemmatization, (v) named entity recognition, (vi) dependency parsing, and (vii) co-reference analysis. These sub-tasks — except task (ii) — were performed using the Stanford CoreNLP suite of tools (CoreNLP, 2014). Sentence splitting and identification of paragraph boundaries was performed using a splitter developed by Piao (2014). Linguistic information extracted by this analysis is exploited by the subsequent parts of the pipeline.

3.2 Identification of Story Characters

The second step is identifying candidate characters (i.e., entities) that appear in the stories under analysis. A story character is not necessarily a story speaker. A character may appear in the story but may not have any quote associated with him and hence, is not a speaker. Characters in children's stories can either be human or non-human entities, i.e., animals and non-living objects, exhibiting anthropomorphic traits. The interactions among characters can either be human-to-human or human-to-non-human interactions.

We used two approaches for identifying story characters motivated by (Elson and McKeown, 2010): 1) named entity recognition was used for identifying proper names, e.g., "Hansel", 2) a set of part-of-speech patterns was used for the extraction of human and non-human characters that were not represented by proper names, e.g., "wolf". The used patterns are: 1) `(DT|CD) (NN|NNS)`, 2) `DT JJ (NN|NNS)`, 3) `NN POS (NN|NNS)`, and 4) `PRP$ JJ (NN|NNS)`.

These POS-based patterns are quite generic, allowing for the creation of large sets of characters. In order to restrict the characters, world knowledge was incorporated through the use of WordNet (Fellbaum, 2005). A similar approach was also followed in (Elson and McKeown, 2010). For each candidate character the hierarchy of its hypernyms was traversed up to the root. Regarding polysemous characters the first two senses were considered. A character was retained if any of its hypernyms was found to fall into certain types of WordNet concepts: person, animal, plant, artifact, spiritual being, physical entity.

3.3 Quote Attribution & Speaker Identification

Here the goal is to attribute (or assign) each quote to a specific story character from the set identified in the previous step. The identification of quotes in the story is based on a simple pattern-based approach: the quote boundaries are signified by the respective symbols, e.g., “ and ”. The pattern is applied at the sentence level.

The quotes are not modeled as NEW/CONT as in (Zhang et al., 2003), however, we adopt a more sophisticated approach for the quote attribution. Three types of attribution are possible in our system: 1) explicit mention of speakers, e.g., “Done!” **said** Hans, merrily, 2) anaphoric mention of speakers, e.g., “How happy am I!” **cried** he, 3) sequence of quotes, e.g., “And where did you get the pig?” ... “I gave a horse for it.”. In the first type of attribution, the speaker is explicitly mentioned in the *vicinity* of the quote. This is also true for the second type, however, a pronominal anaphora is used to refer to the the speaker. The first two attribution types are characterized by the presence of “within-quote” (e.g., “Done!”) and “out-of-quote” (e.g., “said Hans, merrily.”) content. This is not the case for the third attribution type for which only “in-quote” content is available. We refer to such quotes as “pure” quotes. Each attribution type is detailed below.

Preliminary filtering of characters. Before quote-attribution is performed, the list of story characters is pruned by identifying the characters that are “passively” associated with *speech verbs* (SV). This is applied at the sentence level. Some examples of speech verbs are: said, responds, sing, etc. For instance, in “... Hans was **told** ... ”, “Hans” is a passive character. The passive characters were identified via the following relations extracted by dependency parsing: `nsubjpass` (passive nominal subject) and `pobj` (object of a preposition). Given a sentence that includes one or more quotes, the respective passive characters were not considered as candidate speakers. Some other criteria for pruning of list of characters to identify candidate speakers are presented in Section 4.2 (see the three schemes for Tasks 1-2).

Explicit mention of speakers. Several syntactic patterns were applied to associate quotes with explicit mention of speakers in their vicinity to characters from the pruned list of story characters. These patterns were developed around SV. In the example above, “Hans” is associated with the quote “Done!” via the SV “said”. Variations of the following basic patterns (Elson and McKeown, 2010) were used: 1) `QT SV CH`, 2) `QT CH SV`, and 3) `CH SV QT`, where `QT` denotes a quote boundary and `CH` stands for a story character. For example, a variation of the first pattern is `QT SV the? CH`, where `?` stands for zero or one occurrence of “the”.

A limitation of the aforementioned patterns is that they capture associations when the CH and SV occur in close textual distance. As a result, distant associations are missed, e.g., “Hans stood looking on for a while, and at last **said**, “ You must ... ””. In order to address this distant association issue, we examined the `collapsed-ccprocessed-dependencies` output besides the `basic-dependencies` output of the Stanford CoreNLP dependency engine (de Marneffe and Manning, 2012). The former captures more distant relations compared to the latter. We specifically extract the character reference CH either from the dependency relation *nsubj*, which links a speech verb SV with a CH that is the syntactic subject of a clause, or from the dependency relation *dobj*, which links a SV with a CH that is the direct object of the speech verb, across a conjunct (e.g., and). A similar approach was used in (He et al., 2013).

Anaphoric mention of speakers. The same procedure was followed as in the case of the explicit mentions of speakers described above. The difference is that CH included the following pronouns: “he”, “she”, “they”, “himself”, “herself”, and “themselves”. After associating a pronoun with a quote, the quote was attributed to a story character via co-reference resolution. This was done using the co-reference analysis performed by CoreNLP. If a pronominal anaphora was not resolved by the CoreNLP analysis, the following heuristic was adopted. The previous n paragraphs[1] were searched and the pronoun under investigation was mapped to the closest (in terms of textual proximity) story character that had the same gender as the pronoun (see Section 3.4.1 regarding gender estimation). During the paragraph search, anaphoric mentions were also taken into consideration followed by co-reference resolution.

Despite the above approaches, it is possible to have non-attributed quotes. In such cases, the fol-

[1] For the reported results n was set to 5.

lowing procedure is followed for those story sentences that: (i) do not constitute “pure” quotes (i.e., consist of “in-quote” and “out-of-quote” content), and (ii) include at least one “out-of-quote” SV: 1) all the characters (as well as pronouns) that occur within the “out-of-quote” content are aggregated and serve as valid candidates for attribution, 2) if multiple characters and pronouns exist, then they are mapped (if possible) via co-reference resolution in order to narrow down the list of attribution candidates, and 3) the quote is attributed to the nearest quote character (or pronoun). For the computation of the textual distance both quote boundaries (i.e., start and end) are considered. If the quote is attributed to a pronoun that was not mapped to any character, then co-reference resolution is applied.

Sequence of “pure” quotes. Sentences that are “pure” quotes (i.e., include “in-quote” content only) are not attributed to any story character via the last two attribution methods. “Pure” quotes are attributed as follows: The sentences are parsed sequentially starting from the beginning of the story. Each time a character is encountered within a sentence, it is pushed into a “bag-of-characters”. This is done until a non-attributed “pure” quote is found. At this point we assume that the candidate speakers for the current (and next) “pure” quote are included within the “bag-of-characters”. This is based on the hypothesis that the author “introduces” the speakers before their utterances. The subsequent “pure” quotes are examined in order to spot any included characters. Such characters are regarded as “good” candidates enabling the pruning of the list of candidate speakers. The goal is to end up with exactly two candidate speakers for a back and forth dialogue. Then, the initiating speaker is identified by taking into account the order of names mentioned within the quote. Then, the quote attribution is performed in an alternating fashion. For example, consider a sequence of four non-attributed “pure” quotes and a bag of two[2] candidate speakers, s_i and s_j. If s_i was identified as the initiating speaker, then the 1st and the 3th quote are attributed to it, while the 2nd and the 4th quote are attributed to s_j. Finally, the “bag-of-characters” is reset, and the same process is repeated for the rest of the story.

Identification of speakers. The speakers for a given story are identified by selecting those characters that were attributed at least one quote.

3.4 Gender, Age and Personality Attributes

The next three steps in our system involve estimation of the (i) gender, (ii) age, and (iii) personality attributes for the identified speakers.

3.4.1 Gender Estimation

We used a hybrid approach for estimating the gender of the story characters. This is applied to characters (rather than only speakers) because the gender information is exploited during the attribution of quotes (see Section 3.3). The characterization “hybrid” refers to the fusion of two different types of information: (i) linguistic information extracted from the story under analysis, and (ii) information taken from external resources that do not depend on the analyzed story. Regarding the story-specific information, the associations between characters and third person pronouns (identified via anaphora resolution) were counted. The counts were used in order to estimate the gender probability.

The story-independent resources that we used are: (a) the U.S. Social Security Administration baby name database (Security, 2014), in which person names are linked with gender and (b) a large name-gender association list developed using a corpus-based bootstrapping approach, which even included the estimated gender for non-person entities (Bergsma and Lin, 2006). For each entity included in (b) a numerical estimate is provided for each gender. As in the case of story-specific information, those estimates were utilized for computing the gender probability. Using the above information the following procedure was followed for each character: The external resource (a) was used when the character name occurred in it. Otherwise, the information from the external resource (b) and the story-specific information was taken into account. If the speaker was covered by both types of information, the respective gender probabilities were compared and the gender was estimated to be the one corresponding to the highest probability. If the character was not covered by the story-specific information, the external resource (b) was used.

3.4.2 Age Estimation

We used a machine-learning based approach for age estimation. The used features are presented in Table 1, while they were extracted from speaker

[2] If more than two candidates exist, then the system gives ambiguous attributions, i.e., multiple speakers for one quote.

No.	Description
1	count of . , ;
2	count of ,
3	count of !
4	count of 1st person singular pronouns
5	count of negative particles
6	count of numbers
7	count of prepositions
8	count of pronouns
9	count of ?
10	count of tokens longer than 6 letters
11	count of 1st pers. (sing. & plur.) pronouns
12	count of quote tokens
13	count of 1st person plural pronouns
14	count of 2nd person singular pronouns
15	count of quote positive words
16	count of quote negative words
17	count of nouns
18	count of verbs
19	count of adjectives
20	count of adverbs
21	up to 3-grams extracted from quote

Table 1: Common feature set.

quotes, based on the assumption that speakers of different ages use language differently. The development of this feature set was inspired by (Celli, 2012) and (Alm et al., 2005). All features were extracted from the lemmatized form of quotes. Also, all feature counts (except Feature 21) were normalized by Feature 12. For computing the counts of positive and negative words (Feature 15 and 16) we used the General Inquirer database (Stone et al., 1966). Feature 21 stands for n-grams (up to 3-grams) extracted from the speaker quotes. Two different schemes were followed for extracting this feature: (i) using the quote as-is, i.e., its lexical form, and (ii) using the part-of-speech tags of quote. So, two slightly different feature sets were defined: 1) "lex": No.1-20 + lexical form for No.21, 2) "pos": No.1-20 + POS tags for No.21

3.4.3 Estimation of Personality Attributes

A machine-learning based approach was also used for personality attribute estimation. For estimating the personality attributes of story speakers, the linguistic feature set (see Table 1) used in the task for age estimation was used again . Again our approach was based on the assumption that words people use reflect their personality, and the latter can be estimated by these linguistic features.

3.5 Affective Analysis

The last step of our system is the estimation of the affective content of stories. The analysis is performed for each identified quote. The features presented in Table 1 are extracted for each quote and affect is estimated using a machine-learning model, based on the assumption that such features serve as cues for revealing the underlying affective content (Alm et al., 2005; Alm, 2008).

4 Experiments and Evaluation

Here we present the experimental evaluation of our system in performing the following tasks: 1) speaker-to-quote attribution, 2) gender estimation, 3) age estimation, 4) identification of personality attributes, and 5) affective analysis of stories.

4.1 Datasets Used

The datasets used for our experiments along with the related tasks are presented in Table 2.

No.	Task	Type of dataset
1	Quote attribution	STORIES
2	Gender estimation	STORIES
3	Age estimation	QUOTES(1,2)
4	Personality attrib.	QUOTES(3,4)
5	Affective analysis	STORY-AFFECT

Table 2: Experiment datasets and related tasks.

Tasks 1-2. For the first two tasks (quote-to-speaker attribution, and gender estimation) we used a dataset (STORIES) consisting of 17 children stories selected from Project Gutenberg[3]. This set of stories includes 98 unique speakers with 554 quotes assigned to them. The average number of sentences and quotes per story is 61.8 and 32.5, respectively. The average sentence and quote length is 30.4 and 29.0 tokens, respectively. Each speaker was attributed 5.7 quotes on average. Ground truth annotation, which involved assigning quotes to speakers and labeling gender, was performed by one[4] annotator. The following ground truth labels were used to mark gender: "male", "female", and "plural".

[3] www.telecom.tuc.gr/~iosife/chst.html

[4] Due to the limited ambiguity of the task, the availability of a single annotator was considered acceptable.

Task 3. Evaluation of the age estimation task was performed with respect to two different (proprietary) datasets QUOTES1 and QUOTES2. These datasets consisted of individual quotes assigned to popular children's story characters. The dataset QUOTES1 consisted of 6361 quotes assigned to 69 unique speakers. The average quote length equals 7.6 tokens, while each speaker was attributed 141.4 quotes on average. The dataset QUOTES2 consisted of 23605 quotes assigned to 262 unique speakers. The average quote length equals 8.3 tokens, while each speaker was attributed 142.6 quotes on average. For ground truth annotation, four annotators were employed. The annotators were asked to use the following age labels: "child" (0–15 years old), "young adult" (16–35 y.o.), "middle-aged" (36–55 y.o.), and "elderly" (56– y.o.). The age of each character was inferred by the annotators either based on personal knowledge of these stories or by consulting publicly available sources online. The inter-annotator agreement equals to 70%.

Task 4. To evaluate system performance on Task 4, two datasets QUOTES3 and QUOTES4, consisting of individual quotes assigned to popular children's story characters, were used. The set QUOTES3 consisted of 68 individual characters and QUOTES4 consisted of 328 individual characters. The ground truth assignment, assigning each character with personality attributes, was extracted from a free, public collaborative wiki (Wiki, 2014). Since the wiki format allows people to add or edit information, we considered the personality attributes extracted from this wiki to be the average "crowd's opinion" of these characters. Of the open-ended list of attributes that were used to describe the characters, in this task we attempted to extract the following salient personality attributes: "beautiful", "brave", "cowardly", "evil", "feisty", "greedy", "handsome", "kind", "loving", "loyal", "motherly", "optimistic", "spunky", "sweet", and "wise". The pseudo-attribute "none" was used when a character was not described with any of those aforementioned attributes.

Task 5. An annotated dataset, referred to as STORY-AFFECT in this paper, consisting of 176 stories was used. Each story sentence (regardless if quotes were included or not) was annotated regarding primary emotions and mood using the following labels: "angry" (AN), "disgusted" (DI), "fearful" (FE), "happy" (HA), "neutral" (NE), "sad" (SA), "positive surprise" (SU^+), and "negative surprise" (SU^-). Overall, two annotators were employed, while each annotator provided two annotations: one for emotion and one for mood. More details about this dataset are provided in (Alm, 2008).

Instead of using the aforementioned emotions/moods as annotated, we adopted a 3-class scheme for sentence affect (valence): "negative", "neutral", and "positive". In order to align the existing annotations to our three-class scheme the following mapping[5] was adopted: (i) AN, DI, FE, SA were mapped to negative affect, (ii) NE was mapped to neutral affect, and (iii) HA was mapped to positive affect. Given the proposed mapping, we retained those sentences (in total 11018) that exhibited at least 75% annotation agreement.

4.2 Evaluation Results

The evaluation results for the aforementioned tasks are presented below.

Tasks 1-2. The quote-to-speaker attribution was evaluated in terms of precision (AT_p), while the estimation of speakers' gender was evaluated in terms of precision (G_p) and recall (G_r). Note that G_p includes both types of errors: (i) erroneous age estimation, and (ii) estimations for story characters that are not true speakers. In order to exclude the second type of error, the precision of gender estimation was also computed for only the true story speaker identified by the system ($G_p^{'}$). For a subset of the STORIES dataset that included 10 stories, the following schemes were used for filtering of candidate speakers: (i) Scheme 1: all speakers linked with speech verbs, (ii) Scheme 2: speakers, who are persons or animals or spiritual entities according to their first WordNet sense, linked with speech verbs , and (iii) Scheme 3: as Scheme 2,

Speaker filter.	AT_p	G_p	G_r	$G_p^{'}$
Baseline	0.010	0.333		
10 stories (subset of dataset)				
Scheme 1	0.833	0.780	0.672	0.929
Scheme 2	0.868	0.710	0.759	0.917
Scheme 3	0.835	0.710	0.759	0.917
17 stories (full dataset)				
Scheme 2	0.845	0.688	0.733	0.892

Table 3: Quote attribution and gender estimation.

[5] $\text{SU}^{+/-}$ were excluded for simplicity.

but the first two WordNet senses were considered. For the full STORIES dataset (17 stories) Scheme 2 was used. The results are presented in Table 3 including the weighted averages of precision and recall. Using random guesses, the baseline precision is 0.010 and 0.333 for quote-to-speaker attribution and gender estimation, respectively. For the subset of 10 stories, the highest speaker-to-quote attribution attribution is obtained by Scheme 2. When this scheme is applied over the entire dataset, substantially high[6] precision (0.892) is achieved in the estimation of gender of true story speakers.

Task 3. For the estimation of age using quote-based features, a boosting approach was followed using BoosTexter (Schapire and Singer, 2000). For evaluation, 10-fold cross validation (10FCV) was used for the QUOTES1 and QUOTES2 datasets for the "lex" and "pos" feature sets. The results are reported in Table 4 in terms of average classification accuracy. In this table, BOTH refers to the datasets QUOTES1 and QUOTES2 combined together. The evaluation was performed according to two schemes: (i) "relaxed match": the prediction is considered as correct even if it deviates one class from the true one, e.g., "child" and "middle-aged" considered as correct for "young adult", and (ii) "exact match": the prediction should exactly match the true label. The relaxed scheme was motivated by the nature of intended application (storytelling system) for which such errors are tolerable. For the exact match scheme, the obtained performance is higher[7] than the baseline (random guess) that equals to 0.250. The accuracy for the relaxed scheme is quite high, i.e., greater than 0.85 for almost all cases. On average, the "lex" feature set appears to yield slightly higher performance than the "pos" set.

Dataset	Relaxed		Exact	
	lex	pos	lex	pos
Baseline	0.625		0.250	
QUOTES1	0.869	0.883	0.445	0.373
QUOTES2	0.877	0.831	0.450	0.435
BOTH	0.886	0.858	0.464	0.383

Table 4: Age estimation: average accuracy.

Task 4. The personality attributes were estimated using BoosTexter fed with the "lex" feature set. 10FCV was used for evaluation, while the average accuracy was computed by taking into account the top five attributes predicted for each character. The baseline accuracy equals 0.31 given that random guesses are used. Moderate performance was achieved for the QUOTES3 and QUOTES4 datasets, 0.426 and 0.411, respectively.

Task 5. The affect of story sentences was estimated via BoosTexter using the "lex" and "pos" feature sets. As in the previous two tasks 10FCV was applied for evaluation purposes. Using random guesses, the baseline accuracy is 0.33. The average accuracy for the "lex" and "pos" feature sets is 0.838 and 0.658, respectively[8]. It is clear that the use of the "lex" set outperforms the results yielded by the "pos" set.

5 Conclusions and Future Directions

In this paper, we described the development of a multi-step system aimed for story analysis with particular emphasis on analyzing children's stories. The core idea was the integration of several systems into a single pipelined system. The proposed methodology has a strong hybrid character in that it employs different approaches that range from pattern-based to machine learning-based to the incorporation of external knowledge resources. Going beyond the usual task of works in this genre, i.e., speaker-to-quote attribution, the proposed system also supports the estimation of speaker-oriented attributes and affect estimation. Very promising results were obtained for quote attribution and estimation of speaker gender, as well as for age assuming an application-depended error tolerance. The estimation of personality attributes and the affective analysis of story sentences remain open research problems, while the results are more modest especially for the former task.

In the next phase of our work, we hope to improve and generalize each individual component of the proposed system. The most challenging aspects of the system, dealing with personality attributes and affective analysis, will be further investigated. Towards this task, psychological models, e.g., the Big Five model, can provide useful theoretical and empirical findings. Last but not least, the proposed system will be evaluated within the framework of a digital storytelling application including metrics related with user experience.

[6] Statistically significant at 95% lev. (t-test wrt baseline).

[7] Statistically significant at 95% lev. (t-test wrt baseline).

[8] Statistically significant at 90% lev. (t-test wrt baseline).

References

C. O. Alm, D. Roth, and R. Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proc. of Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 579–586.

C. O. Alm. 2008. *Affect in Text and Speech*. Ph.D. thesis, University of Illinois at Urbana-Champaign.

K. Balog, G. Mishne, and M. de Rijke. 2006. Why are they excited? identifying and explaining spikes in blog mood levels. In *Proc. 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 207–210.

S. Bergsma and D. Lin. 2006. Bootstrapping path-based pronoun resolution. In *Proc. of Conference on Computational Lingustics / Association for Computational Linguistics*, pages 33–40.

J. Cabral, L. Oliveira, G. Raimundo, and A. Paiva. 2006. What voice do we expect from a synthetic character? In *Proceedings of SPECOM*, pages 536–539.

F. Celli. 2012. Unsupervised personality recognition for social network sites. In *Proc. of Sixth International Conference on Digital Society*.

CoreNLP. 2014. Stanford CoreNLP tool. `http://nlp.stanford.edu/software/corenlp.shtml`.

M.-C. de Marneffe and C. D. Manning. 2012. Stanford typed dependencies manual.

D. K. Elson and K. R. McKeown. 2010. Automatic attribution of quoted speech in literary narrative. In *Proc. of Twenty-Fourth AAAI Conference on Artificial Intelligence*.

C. Fellbaum. 2005. Wordnet and wordnets. In K. Brown et al., editor, *Encyclopedia of Language and Linguistics*, pages 665–670. Oxford: Elsevier.

E. Greene, T. Mishra, P. Haffner, and A. Conkie. 2012. Predicting character-appropriate voices for a TTS-based storyteller system. In *Proc. of Interspeech*.

H. He, D. Barbosa, and G. Kondrak. 2013. Identification of speakers in novels. In *Proc. of 51st Annual Meeting of the Association for Computational Linguistics*, pages 1312–1320.

M. Hu and B. Liu. 2004. Mining and summarizing customer reviews. In *Proc. of Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177.

L. Lloyd, D. Kechagias, and S. Skiena. 2005. Lydia: A system for large-scale news analysis. In *Proc. SPIRE*, number 3772 in Lecture Notes in Computer Science, pages 161–166.

N. Malandrakis, A. Potamianos, E. Iosif, and S. Narayanan. 2013. Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11):2379–2392.

N. Mamede and P. Chaleira. 2004. Character identification in children stories. In J. Vicedo, P. Martnez-Barco, R. Muoz, and M. Saiz Noeda, editors, *Advances in Natural Language Processing*, volume 3230 of *Lecture Notes in Computer Science*, pages 82–90. Springer Berlin Heidelberg.

R. Mitkov. 2002. *Anaphora Resolution*. Longman.

I. R. Murray and J. L. Arnott. 2008. Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. *Computer Speech and Language*, 22(2):107–129.

P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson. 2013. Semeval 2013 task 2: Sentiment analysis in twitter. In *Proc. of Second Joint Conference on Lexical and Computational Semantics (*SEM), Seventh International Workshop on Semantic Evaluation*, pages 312–320.

T. Natsiopoulou, M. Souliotis, and A. G. Kyridis. 2006. Narrating and reading folktales and picture books: storytelling techniques and approaches with preschool children. *Early Childhood Research and Practice*, 8(1). Retrieved on Jan 13th, 2014 from `http://ecrp.uiuc.edu/v8n1/natsiopoulou.html`.

T. W. Norman. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. *Journal of Abnormal and Social Psychology*, 66:574–583.

S. Piao. 2014. Sentence splitting program. `http://text0.mib.man.ac.uk:8080/scottpiao/sent_detector`.

D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta. 2010. Classifying latent user attributes in twitter. In *Proc. of the 2nd International Workshop on Search and Mining User-generated Contents*, pages 37–44.

M. Rusko, M. Trnka, S. Darjaa, and J. Hamar. 2013. The dramatic piece reader for the blind and visually impaired. In *Proc. of 4th Workshop on Speech and Language Processing for Assistive Technologies*, pages 83–91.

R. E. Schapire and Y. Singer. 2000. Boostexter: A boosting-based system for text categorization. *Machine. Learning*, 39(2-3):135–168.

J. Schler, M. Koppel, S. Argamon, and J. W. Pennebaker. 2006. Effects of age and gender on blogging. In *Proc. of AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*.

Social Security. 2014. U.S. social security administration baby name database. `http://www.ssa.gov/OACT/babynames/limits.html`.

P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

C. Strapparava and R. Mihalcea. 2007. Semeval 2007 task 14: Affective text. In *Proc. SemEval*, pages 70–74.

M. Theune, K. Meijs, and D. Heylen. 2006. Generating expressive speech for storytelling applications. In *IEEE Transactions on Audio, Speech and Language Processing*, pages 1137–1144.

P. Turney and M. L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus (technical report erc-1094).

Disney Wiki. 2014. Description of Disney characters. `http://disney.wikia.com/wiki/Category:Disney_characters#`.

J. Y. Zhang, A. W. Black, and R. Sproat. 2003. Identifying speakers in children's stories for speech synthesis. In *Proc. of Interspeech*.