Identifying Principals and Accessories in a Complex Case based on the Comprehension of Fact Description

[1]**Yakun Hu**[*], [2]**Zhunchen Luo**[*], [1]**Wenhan Chao**[*†]

[1]School of Computer Science and Engineering, Beihang University
[2]Information Research Center of Military
PLA Academy of Military Science
[1,2]Beijing, China
[1]{huyakun,chaowenhan}@buaa.edu.cn
[2]zhunchenluo@gmail.com

Abstract

In this paper, we study the problem of identifying the principals and accessories from the fact description with multiple defendants in a criminal case. We treat the fact descriptions as narrative texts and the defendants as roles over the narrative story. We propose to model the defendants with *behavioral semantic information* and *statistical characteristics*, then learning the importances of defendants within a learning-to-rank framework. Experimental results on a real-world dataset demonstrate the behavior analysis can effectively model the defendants' impacts in a complex case.

1 Introduction

In recent years, much previous work has focused on the building of legal assistant systems with different functions, e.g. searching relevant cases for a given query (Chen et al., 2013), predicting charge labels based on the fact description in a criminal case (Luo et al., 2017; Hu et al., 2018; Zhong et al., 2018), generating the interpretable court views from the fact descriptions (Ye et al., 2018). Though having achieved promising results in this field, most of the work only studies the simple cases with only one defendant. However, there exist lots of complex criminal cases in practice, which will involve multiple criminals.

In this work, *we propose to study the identification of principals and accessories from the fact description in a criminal case*. The principal refers to a criminal who organizes and leads criminal groups to carry out criminal activities or plays a major role in joint crimes. Correspondingly, we refer the accessory as the one who plays a secondary or auxiliary role. As the illustration of our task in Fig. 1, given the fact description as well as a list of defendants, we expect to identify the principals and accessories from the defendants.

Since the fact descriptions in criminal cases are usually narrative texts which mostly record the criminal events, we treat the defendants in the fact descriptions as the narrative roles and the protagonists who have the greater impact will be identified as the principles. Narrative comprehension has been studied in NLP for a long time. The traditional method to measure the importances of roles is based on the roles' dispersion over the story (Karsdorp et al., 2012). It supposes that comparing to less important roles, the roles with bigger impact are expected to appear at more places and are more evenly distributed over the story. However, this assumption ignores actions of roles (denoted as *behavioral semantic information*), which may be a key factor that estimates their impacts in legal-context scenarios. In this paper, we propose to model a defendant from two perspectives of *behavioral semantic information* and *statistical characteristics*. After that, we further learned to estimate their importances with a learning-to-rank framework (Joachims et al., 2007). Our contributions in this paper can be summarized as:

• We are the first to identify principals and accessories from complex cases with multiple defendants based on the comprehension of a narrative fact description.

• We treat the fact descriptions as narrative texts and the defendants as roles in a narrative story.

• We propose to model a defendant with *semantic information* and *statistical characteristics* and estimate his importance within a learning-to-rank framework.

2 Related Work

Our work is a task related to narrative comprehension. There has recently been a upsurge in re-

---

[*] indicates equal contribution.
[†] Corresponding author.

Fact Description:
[谢军波先起意抢劫，经得仇某同意后，俩人尾随张某至本市大溪镇闸头村桥头边，谢军波上前将张某推倒在地后，接着抢走一只挎包，包内有一部诺基亚手机、现金人民币 10 多元、一张身份证、一张银行卡等物。]#[Xie Junbo started to rob. With the consent of Qiu, they followed Zhang to the bridgehead of Zhatou Village in Daxi Tawn. Xie Junbo pushed Zhang to the ground and then stole a bag, which contained a Nokia mobile phone, more than 10 yuan in cash, and a ID card, a bank card and so on.]
Defendants:
[谢军波, 仇某]#[Xie Junbo, Qiu]
Principal: [谢军波]#[Xie Junbo]
Accessory: [仇某]#[Qiu]

Figure 1: An example of a case involving two defendants.

search in information extraction of narrative and story understanding. Ouyang and McKeown (2015) presents a change-based model to capture the rise and fall of story characteristics within narrative. Goh et al. (2012) proposes to identity the protagonist in fairy tales automatically with the aid of verbs. Karsdorp et al. (2012) presents a method for extraction the cast from fictional texts and ranks the different cast members on a scale of importance to the story on the basis of their dispersion in the text. However, it only considers the position information and ignores the behavioral semantic information.

Meanwhile, the task is related to the researches on legal assistant system. Studies on the application of machine learning in the judicial field have been concentrated in the following directions: learning to predict charges for criminal cases given the fact descriptions (Luo et al., 2017; Jiang et al., 2018; Chao et al., 2019), identifying applicable articles for a given case (Liu and Liao, 2005), providing a tool for automated text summary of legal documents based on word frequency augmented with additional domain-specific knowledge (Polsley et al., 2016). In addition, Ye et al. (2018) put forward a new task of COURT-VIEW-GEN that generates court view from the fact description. But those studies do not involve complex cases involving multiple criminals.

3 Methodology

Given the fact description f of a case and its defendants set $d(d_1, d_2, ..., d_n)$, we expect to classify each d_i as either a principal or an accessory. A function F for scoring each defendant is learned by a ranking method and we regard its result as the probability of d_i being a principal. Note that this could be treated as a classification problem without loss of generality.

3.1 Features

We consider two feature families when modeling a defendant: behavioral semantic features (denoted as **f_semantic**, including *Activity Fragments*) and statistical characteristics (denoted as **f_statistical**, including *Sentence Syntactic Complexity*, *Cooperation Mode* and *Order and Frequency*).

Activity Fragments: Sentences containing one's actions can reflect his impact in the case to a great extent. Then, we select sentences by name for each defendant and filter out those without verbs. We feed them and the total fact description into two bidirectional *LSTMs* (Schuster and Paliwal, 1997) for automatic semantic information extraction. Next, we introduce *match-lstm* (Wang and Jiang, 2016) to fuse those two outputs to measure a defendant's impact on the case. The output from total fact description corresponds to the hypothesis of *match-lstm* and that from activity fragments corresponds to the premise. Finally, the output of the *match-lstm* is treated as the behavioral semantic features (**f_semantic**) of a defendant.

Sentence Syntactic Complexity: The principal of a case is defined as a person who plays a major role in criminal activities. Then, he may appear in more sentences and there may be more verbs related to him. Accordingly, we utilize the syntactic complexity of sentences (Ouyang and McKeown, 2015) as an important feature. Several statistical characteristics are considered to model the syntactic complexity, including the length of the sentence (**sentlength**), the length of its verb phrases (**vplength**), the depth of the sentence's parse tree (**sentdepth**) (Klein and Manning, 2003), the depth of the verb phrase's parse tree (**vpdepth**), average number of words (**avgwords**), average number of verds (**avgverbs**).

Cooperation Mode: Moreover, the protagonist is often the plotter of the story and it can be expressed as who is the planner of the case. This information is often reflected by some verbs or con-

经事先联系，申春西陪同杨惠军至本市城北街道方家村余某的出租房屋内…

After contacting in advance, Shen Chunxi accompanied Yang Huijun to the Yu's lodge in Fangjia village, Chengbei street…

2013 年 9 月至 11 月期间，被告人陶某、吴某、高某甲经事先预谋…

During the period from September to November 2017, the defendants Tao, Wu and Gao were premeditated…

Figure 2: Two example that reflects the defendants' co-operation mode in a case. The word in red represents a master-slave relation between two defendants and the word in green represents equality relation.

junctions and can be obtained by mining the **cooperation_mode** (could be master-slave or equality relation) between defendants as shown in Fig. 2. We construct a verb set and a conjunction set that could reflect the cooperation mode manually from our corpus of criminal cases. Then we utilize the Stanford CoreNLP (Manning et al., 2014) to find out the conjunction or verb between two defendants. Finally, it is mapped to a vector based on which set of cooperation mode it belongs to.

Order and Frequency: Finally, we propose two other potentially useful features. One is the **order_of_appearances** of the defendants and the other is the **number_of_occurrences**. We suppose the principal of a case is the plotter and naturally should appears in the fact description earlier. Besides, defendant with more frequent occurrences probably has a greater impact on the case.

3.2 Ranking Model

We utilize RankNet (Burges et al., 2005) to train our ranking model. We calculate scores for both **f_semantic** and **f_statistical** respectively and regard their weighted sum as the final probability of being a principal. The scoring units are all composed of linear functions.

4 Experiments

4.1 Data Preparation

Extensive experiments are conducted on a real world dataset obtained from Chinese government website[*] to evaluate our method. Following Ye et al. (2018), we regard the paragraph start with "our court identified that" and end with "the above facts" as the fact description. Burges et al. (2005) shows that training on ties makes little difference. Therefore, we could consider only defendant pairs (A, B) such that A plays a more important role than

---
[*] http:/wenshu.court.gov.cn

B and label it 1. Accordingly, samples containing only principals or accessories are labelled 0. We get a total of 15312 criminal cases with more than two defendants and the percentage of cases involving only one principal is 67%. Finally, 41342 paired samples are generated. Summary statistics of the data are listed in Tab. 1. To verify the reliability and stability of the model, we perform 10-fold cross-validation in our dataset.

total cases	cases@2	cases@3	cases@4+
15312	7364	5016	2932

Table 1: Summary statistics of the data. **cases@n** referrings to the number of cases with n defendants and "+"means the number is not less than n.

4.2 Settings

The dimension of word embedding is 200 and dimension of hidden states in BiLstm is set to 256. In addition, mini-batch size is set to 32 and the default learning rate of Adam (Kingma and Ba, 2014) is $1e - 3$.

4.3 Baselines

Previous studies on the importance distinction of roles in narrative texts are mainly based on statistical features and we are merely exploring solutions to this new problem proposed by this paper. Our baselines are as follows:

- *Frequency*: A basic method in which retrieved items are ranked according to their number of occurrences.

- *Dispersion*: The basic idea is that more important roles are expected to appear at more places in the story and are more eventually distributed over the story than less important roles (Karsdorp et al., 2012).

- *Frequency&Dispersion*: We combine the two methods above as our third baseline.

Model	P_{macro}	R_{macro}	F_{macro}
Frequency	66.54	63.73	65.10
Dispersion	71.28	69.44	70.35
Frequency&Dispersion	74.15	72.34	73.23
Ours	**80.36**	**79.18**	**79.77**

Table 2: The performances of different role modeling methods.

5 Results and Discussion

Tab. 2 presents the performances of different role modeling methods. It can be seen that our model achieves a considerable improvement in P_{macro}, R_{macro} and F_{macro}. As shown in Fig. 3, the defendant in red is the mastermind of the case and should be judged to be the principal, despite his low appearances. Position or frequency information does not effectively reflect the status of a role in such samples. However, our method captures this information by the cooperation mode feature between *Yin* and *Zhao*, with the help of verb "*instructed*".

尹某指使赵某甲做卖毒人，一起将 2 小包冰毒以 1000 元的价格卖给王朋，由被告人赵某甲收取了毒资人民币 1000 元。

Yin instructed Zhao to be a drug seller and sold two small packages of methamphetamine to Wang Peng at the price of 1000 yuan. Zhao collected 1000 yuan of poison money.

Figure 3: A case in where defendant having low occurrence frequency is convicted as a principal.

We compare the performances of our two feature families to explore which one contributes more to the task and the result is shown in Tab. 4. We find that the feature family **f_semantic** achieve better performance in all the evaluation metrics. And its results are even better than our baselines. It reveals that defendant's behavioral semantic information is more valuable than those statistical characteristics.

Feature Family	P_{macro}	R_{macro}	F_{macro}
f_statistical	75.18	72.79	73.97
f_semantic	**77.26**	**74.64**	**75.93**

Table 3: The performances of different feature families.

6 Feature Selection

We expect to find a feature conjunction that makes the most sense for modeling role's impact in a story. Like Duan et al. (2010), we use an advanced greedy method to find the best feature conjunction. Given all n (it is 10 in this paper) features we extracted, we construct 2^n feature sets and randomly pick 100 of them. Then, we run the greedy selection algorithm based on the feature set (denoted as **Best**) with the best *MAP* among those 100 feature sets. Features excluded those in **Best** are denoted as **Ex_best** and all the extracted features are denoted as **Full**. We evaluate the **Best** and each feature in **Ex_best** and if the result is better than the previous one, this feature will be added into the **Best**. We repeat the process until the **Best** is no longer updated. Finally, we get the best feature conjunction composed by **f_semantic**, **vpdepth**, **order_of_appearances**, **number_of_occurrences**, **cooperation_mode**. To reflect the gap between the **Best** and the **Full**, we evaluate their performances on datasets with different numbers of defendants. Tab. 4 illustrates the **Best** feature set also outperforms the **Full** feature set when dealing with cases with different numbers of defendants.

Model	F_{macro} (%)		
	$\#Def = 2$	$\#Def = 3$	$\#Def = 4^+$
Full	81.33	80.62	77.28
Best	82.85	81.41	79.54

Table 4: $F_{macro}(\%)$ of **Best** and **Full** on datasets with different numbers of defendants (denoted as $\#Def$).

We are interested in which features in particular are highly valued for role modeling. The importance of each feature is evaluated by the decrease of performance when removing this feature measured from the **Best**. Fig. 4 reveals the importance of each feature for role modeling.
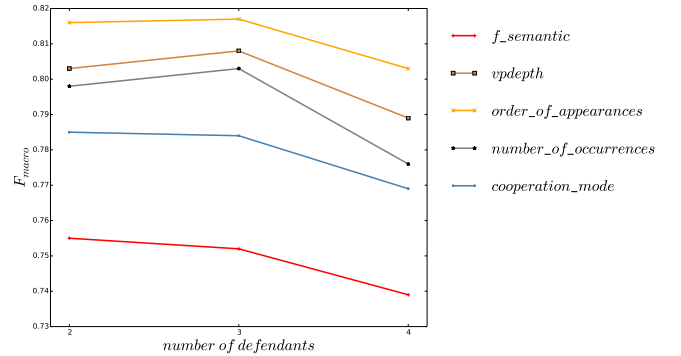

Figure 4: Importance of each feature

We observe that **f_semantic** plays a very important role. The F_{macro} declines seriously (more than 6 percentage points) when we remove it from the feature set. We suppose that semantic features represent the behavioral information of roles and a defendant's behavior is of great concern in determining his criminal responsibility. The match result of a defendant's actions and global description of the case can effectively model his influence in the whole case.

7 Conclusion

In this paper, we study the task of identifying principals and accessories from the fact description in a complex case. We find a set of effective features for role modeling. and evaluate that the behavioral semantic information is most worthy of attention. We hope to address this problem with a completely semantic-based approach in the future.

Acknowledgments

References

Christopher J. C. Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Gregory N. Hullender. 2005. Learning to rank using gradient descent. In *ICML*.

Wenhan Chao, Xin Jiang, Zhunchen Luo, Yakun Hu, and Wenjia Ma. 2019. Interpretable charge prediction for criminal cases with dynamic rationale attention. *J. Artif. Intell. Res.*, 66:743–764.

Yen-Liang Chen, Yi-Hung Liu, and Wu-Liang Ho. 2013. A text mining approach to assist the general public in the retrieval of legal documents. *JASIST*, 64(2):280–290.

Yajuan Duan, Long Jiang, Tao Qin, Ming Zhou, and Heung-Yeung Shum. 2010. An empirical study on learning to rank of tweets. In *COLING*.

Hui-Ngo Goh, Lay-Ki Soon, and Su-Cheng Haw. 2012. Automatic identification of protagonist in fairy tales using verb. In *PAKDD (2)*.

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2018. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 487–498. Association for Computational Linguistics.

Xin Jiang, Hai Ye, Zhunchen Luo, Wenhan Chao, and Wenjia Ma. 2018. Interpretable rationale augmented charge prediction system. In *COLING 2018, The 27th International Conference on Computational Linguistics: System Demonstrations, Santa Fe, New Mexico, August 20-26, 2018*, pages 146–151. Association for Computational Linguistics.

Thorsten Joachims, Hang Li, Tie-Yan Liu, and ChengXiang Zhai. 2007. Learning to rank for information retrieval (LR4IR 2007). *SIGIR Forum*, 41(2):58–62.

Folgert Karsdorp, Peter Van Kranenburg, Theo Meder, and Antal Van Den Bosch. 2012. Casting a spell: Identification and ranking of actors in folktales. In *The Workshop on Annotation of Corpora for Research in Het Humanities*.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *ACL*.

Chao-Lin Liu and Ting-Ming Liao. 2005. Classifying criminal charges in chinese for web-based legal services. In *APWeb*.

Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to predict charges for criminal cases with legal basis. In *EMNLP*.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David Mc-Closky. 2014. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*.

Jessica Ouyang and Kathleen R. McKeown. 2015. Modeling reportable events as turning points in narrative. In *EMNLP*.

Seth Polsley, Pooja Jhunjhunwala, and Ruihong Huang. 2016. Casesummarizer: A system for automated summarization of legal texts. In *COLING (Demos)*.

Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Processing*, 45(11):2673–2681.

Shuohang Wang and Jing Jiang. 2016. Learning natural language inference with LSTM. In *HLT-NAACL*.

Hai Ye, Xin Jiang, Zhunchen Luo, and Wenhan Chao. 2018. Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. In *NAACL-HLT*.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3540–3549. Association for Computational Linguistics.