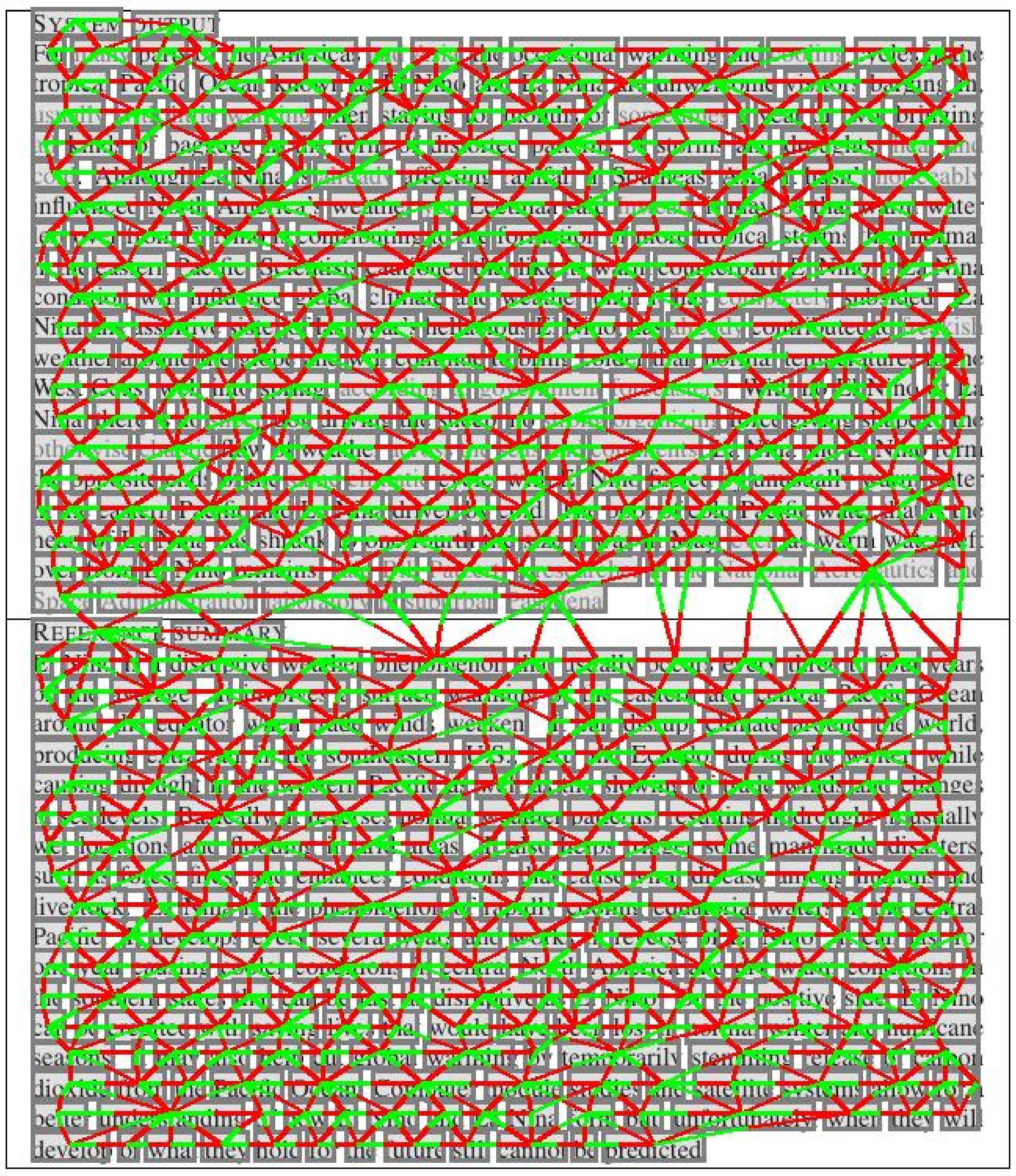
Table 2: Human evaluation results

Written	GR	NR	RC	TF	SC
SpOpt-Δ	4.20±0.70	3.53±0.62	3.93±0.93	3.80±0.87	3.40±0.66
SpOpt-comp-Δ	3.43±0.84	3.97±0.60	4.00±0.86	3.90±0.79	3.73±0.68
Human	4.97±0.18	4.93±0.25	5.00±0.00	4.93±0.25	4.93±0.25

Table 3: Example compressive summary from DUC 2006



5 Related Work

Our sparse optimization formulations are closely related to data reconstruction for document summarization. The data reconstruction paradigm for document summarization, originally proposed by He et al. [2012], was inspired by latent semantic analysis (LSA) that utilizes singular value decomposition (SVD) to select highly ranked sentences [Gong and Liu, 2001]. Nonnegative matrix factorization has also been introduced to group sentences into clusters [Wang *et al.*, 2008]. Recently Liu et al. [2015] propose a two-level sparse representation model. Their optimization problem is NP-hard so heuristic methods such as simulated annealing has been used to solve it approximately.

In recent years some research has made much progress beyond extractive summarization, especially in the context of compressive summarization. An earlier attempt made by Zajic et al. [2006] tried a pipeline strategy with heuristics to generate multiple candidate compressions and extract from this compressed sentences. Berg-Kirkpatrik et al. [2011] created linear models of weights learned by structural SVMs for different components and tried to jointly do sentence selection and syntax tree trimming in integer linear programs. Woodsend and Lapata [2012] designed quasi tree substitution grammars for multiple rewriting operations. All these methods involve integer linear programming solvers to generate the final compressed summary, which is time-consuming for multi-document summarization tasks.

Almeida and Martins [2013] formed the compressive summarization problem in a more efficient dual decomposition framework. Models for sentence compression and extractive summarization are trained by multi-task learning techniques. Wang et al. [2013] explored different types of compression on constituent parse trees for query-focused summarization. In these works, the best-performing systems require supervised learning for different subtasks.

Our mathematical formulations are closely related to modern sparse optimization problems. Subspace clustering techniques [Elhamifar and Vidal, 2013] try to learn proper coefficients, aiming at a self-representation. The difference between general sparse subspace clustering problems and our formulations will make key impact on the choice and design of solving algorithms. The difference comes mainly from different motivations. The former expect for sparsity in sentence selection, while the latter typically requires low-rankness in matrix representations.

6 Conclusion and Future Work

In this paper we propose a new formulation for document summarization via sparse optimization with decomposable convex objective function and derive an efficient ADMM algorithm to solve it. We also introduce a sentence dissimilarity term to encourage diversity in summaries. Then we generalize the proposed method to compressive summarization and derive a block coordinate descent procedure along with recursive dependency tree compression to generate the final sentences. Experimental study shows that our compressive summarization framework significantly improves results from the original extractive methods based on data reconstruction.

Structured sparsity has been studied for a while in machine learning community. However, its adaptation in natural language processing and text mining is still at its beginning [Martins *et al.*, 2011; Yogatama and Smith, 2014]. We would like to explore if structured sparsity can become useful for compressive summarization tasks.

Our proposed methods are fully unsupervised. We would like to extend it to supervised case for different optimization problems described in this paper. It is reasonable to expect for even better performance.

Acknowledgments

We thank all the anonymous reviewers for helpful comments and suggestions. This work was supported by National Hi-