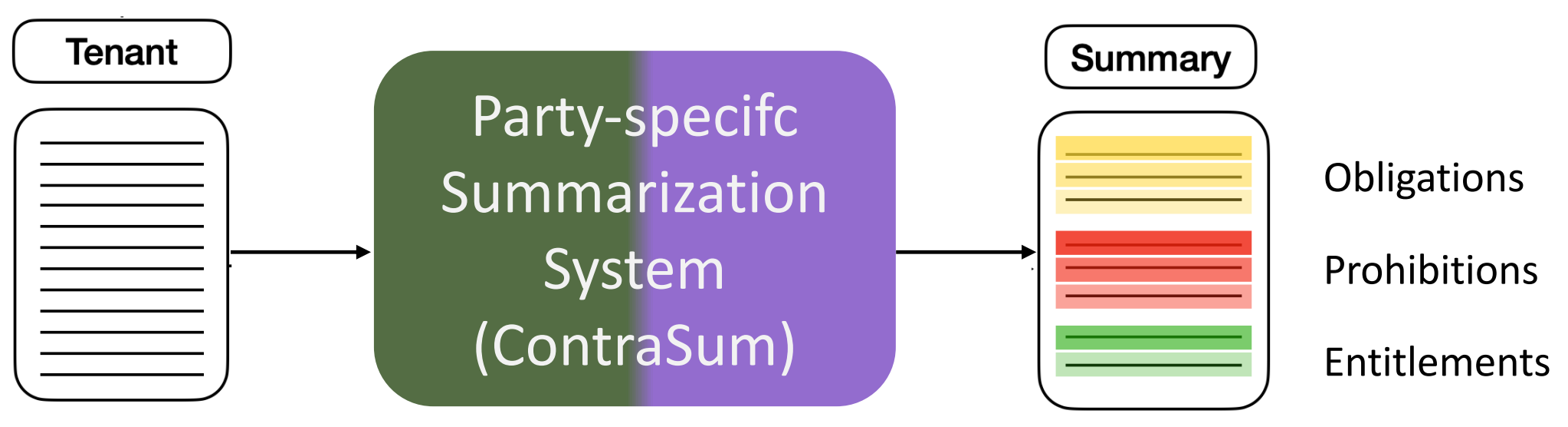
What to Read in a Contract? Party-specific Summarization of Legal Obligations, Entitlements, and Prohibitions

Abhilasha Sancheti[1,2], Aparna Garimella[2], Balaji Vasan Srinivasan[2], Rachel Rudinger[1]
[1] *University of Maryland, College Park* [2] *Adobe Research*
sancheti@umd.edu, garimell@adobe.com, balsrini@adobe.com, rudinger@umd.edu

1. Information Overload

- Existing works [1, 2, 3] help in extracting relevant information
  - 50+ sentences per category (obligations, entitlement, prohibitions)
- **Solution: Summarize a contract?**
  - single summary may not serve all the parties as they have different rights and duties
  - all the obligations (or other categories) are **not equally important** (*e.g.*, higher liability obligations are more important than others for a party)

RQ: How can we automatically generate an "at a glance" summary of *important* rights and duties for each contracting party?



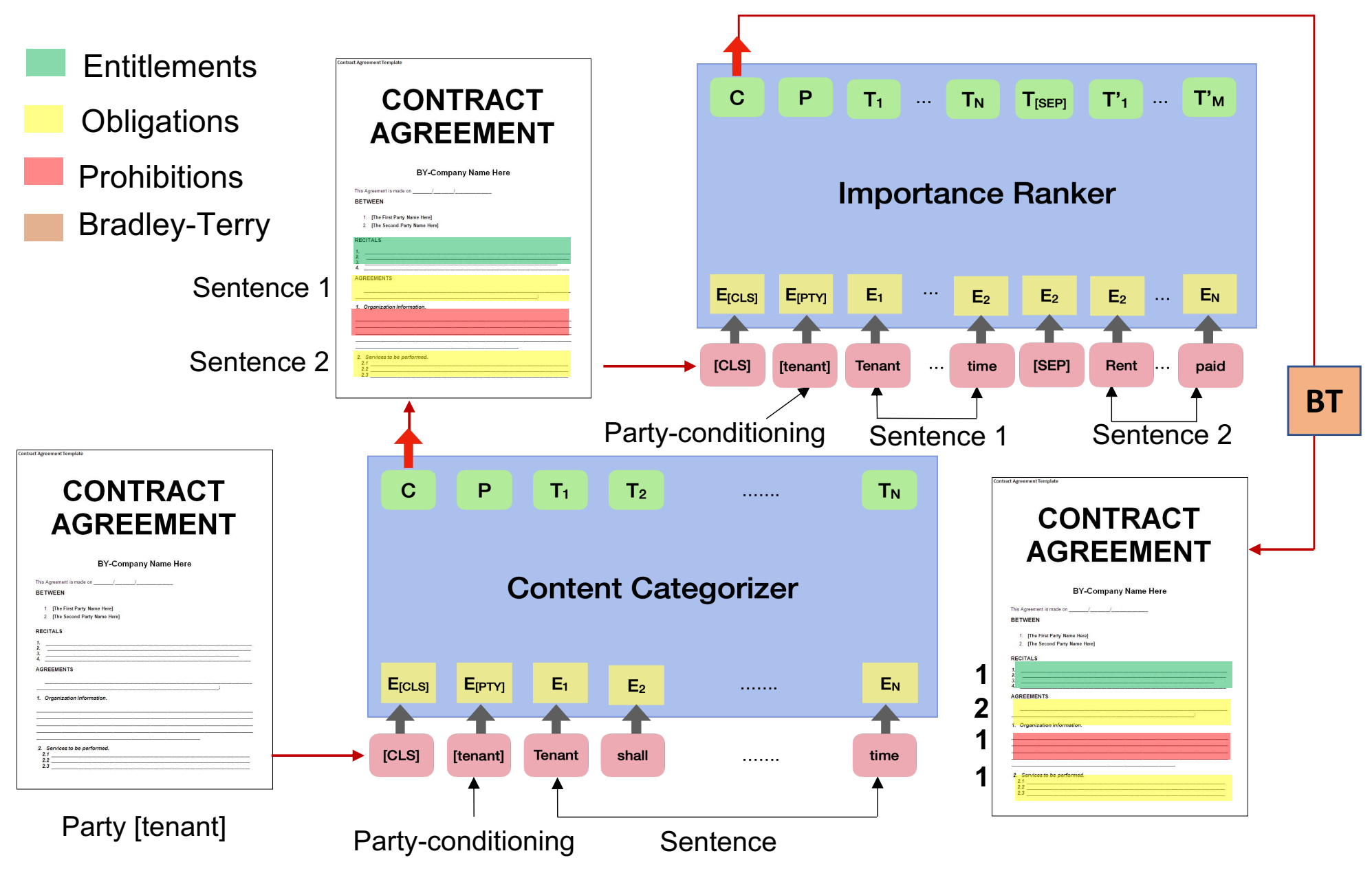
2. Collecting *Importance* annotations is challenging

- **Extend** the LexDeMod dataset [3] with *party-specific importance annotations*
- **Problem: Rating importance of a sentence on a scale**
  - requires **well-defined levels**; can be **subjective, and restrictive**
  - prone to **difficulty in maintaining inter- and intra-annotator consistency** [4]
  - low inter-annotator agreement in pilot studies for rating single sentence and pair of sentences
- **Solution:** Best-worst scaling [5]

Annotation Task: (Party, S_1, S_2, S_3, S_4) Most important? Least important?
S_i = sentences containing obligations, entitlements, prohibitions for a Party from LexDeMod dataset

- Do not provide a detailed technical definition for *importance* instead
  - **brief legal experts** about the **task of summarization** from **review and compliance**'s perspective
  - **encourage** them **to rely** on their **intuition, experience, and expertise**
- ∼293K pairwise importance comparisons; Moderate-high reliability (SHR=0.66 ± 0.01)
- Prohibitions > Obligations > Entitlements for **Tenant** (*e.g.*, severe penalties associated with prohibitions)
- Entitlements > Obligations > Prohibitions for **Landlord** (*e.g.*, landlords face fewer prohibitions and obligations than tenants)

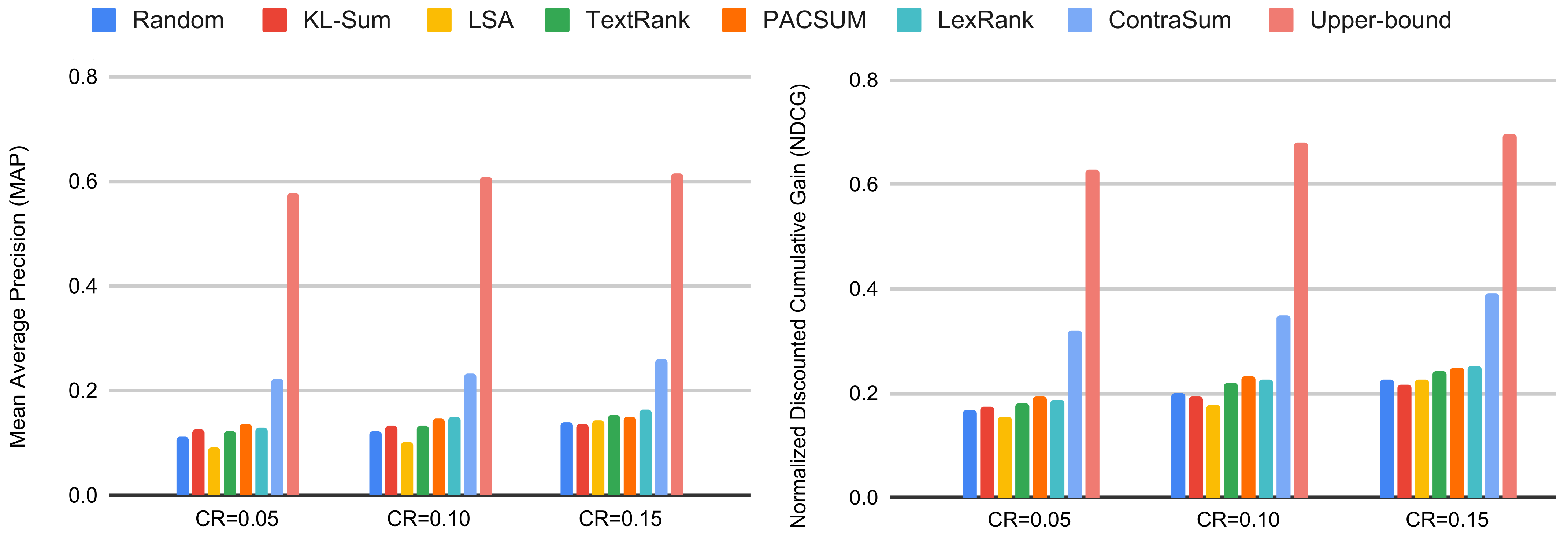
3. ContraSum



ContraSum takes a contract and a party to first identify all the sentences containing *party-specific* obligations, entitlements, and prohibitions using a **content categorizer**. Then, the identified sentences within each category are pairwise importance-ranked for a given party using an **importance ranker**. A full ranked list of sentences is obtained using the Bradley-Terry model to obtain the final summary.

4. Automatic Evaluation

Dataset: Contracts from LexDeMod with category + importance annotations



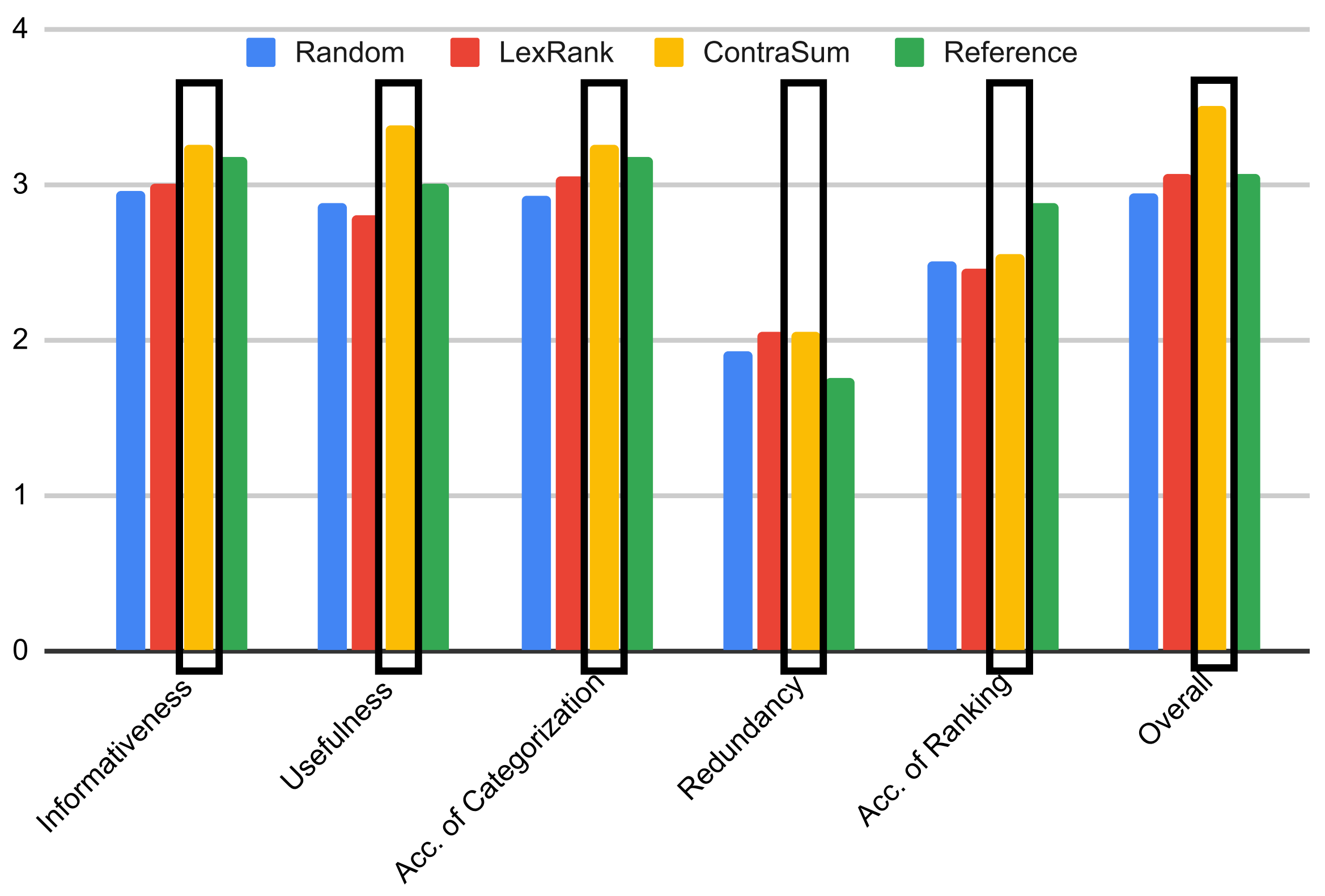
Same predicted categories (similar trends with ground-truth categories) are used by all the systems

Takeaway:
1. ContraSum obtains the highest MAP and NDCG scores establishing the need for modeling domain-specific notion of importance.
2. Huge gap between ContraSum and Upper-bound indicates modeling *party-specific* importance comparison is a hard task.

5. Legal Expert Evaluation

- 2 legal experts rate summaries for each party
- Max. 10 sentences per category per summary
- Rate the summaries on 5-point scale per category per party for 5 criteria: **Informativeness, Usefulness, Accuracy of categorization, Accuracy of importance ranking, Redundancy**
- **Overall:** quality of overall summary?

Takeaway: Summaries from ContraSum are informative, useful, and correctly categorized



6. Conclusion

- Introduce a new task, dataset, and a system for *party-specific* summarization of contracts
- Breaking the task of summarization into two sub-tasks of categorization and ranking that enables
  - use of existing categorization dataset
  - development of ContraSum; needs much less data than an end-to-end supervised summarization system

Paper
Dataset

References

[1] Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. Obligation and prohibition extraction using hierarchical rnns. In *ACL 2018.*

[2] Elliott Ash, Jeff Jacobs, Bentley MacLeod, Suresh Naidu, and Dominik Stammbach. Unsupervised extraction of workplace rights and duties from collective bargaining agreements. In *2020 ICDMw*. IEEE.

[3] Abhilasha Sancheti, Aparna Garimella, Balaji Vasan Srinivasan, and Rachel Rudinger. Agent-specific deontic modality detection in legal language. In *Proceedings of the 2022 Conference on EMNLP.*

[4] Svetlana Kiritchenko and Saif M Mohammad. Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. In *Proceedings of the 2017 Conference on ACL.*

[5] Jordan J Louviere and George G Woodworth. Best-worst scaling: A model for the largest difference judgments. Technical report, Working paper, 1991.

Acknowledgements

This work was supported by Adobe Research unrestricted gift funding.