

ECO v1: 以事件为中心的观点挖掘

Findings of the Association for Computational Linguistics: ACL 2022. 2022: 2743-2753

徐若曦^{1,3}, 林鸿宇^{1*}, 廖梦^{4*}, 韩先培^{1,2}, 徐进⁴, 谭维⁴, 孙应飞³, 孙乐^{1,2}¹中文信息处理实验室 ²中国科学院软件研究所计算机科学国家重点实验室³中国科学院大学 ⁴腾讯微信数据质量组

背景

- 传统的观点挖掘任务以实体为中心，将事件视作一种特殊实体类型；
- 缺少以事件为中心的观点挖掘数据集。

实体: 佳能相机

这款相机很坚固, 拍照也不错, 如果你关闭LCD取景器, 电池也可以用很长时间。

(佳能相机, GENERAL, 坚固, 正面)
(佳能相机, 拍照, 不错, 正面)
(佳能相机, 电池, 用很长时间, 正面)

以实体 VS. 事件为中心的观点挖掘

- 以事件为中心的观点挖掘任务与传统的以实体为中心的观点挖掘任务在定义、结构、表达和分布方面有显著不同。

	以实体为中心	以事件为中心
定义	情感极性	判断、态度、情感信仰、建议
结构	实体及其属性	事件及其子事件、相关事件、参与者
表达	更多显式	更多隐式
分布	简短而集中的评论	更长的新闻或文档

任务定义

以事件为中心的观点挖掘任务旨在分别从给定的文档和事件短语中提取 (观点, 论元) 对。

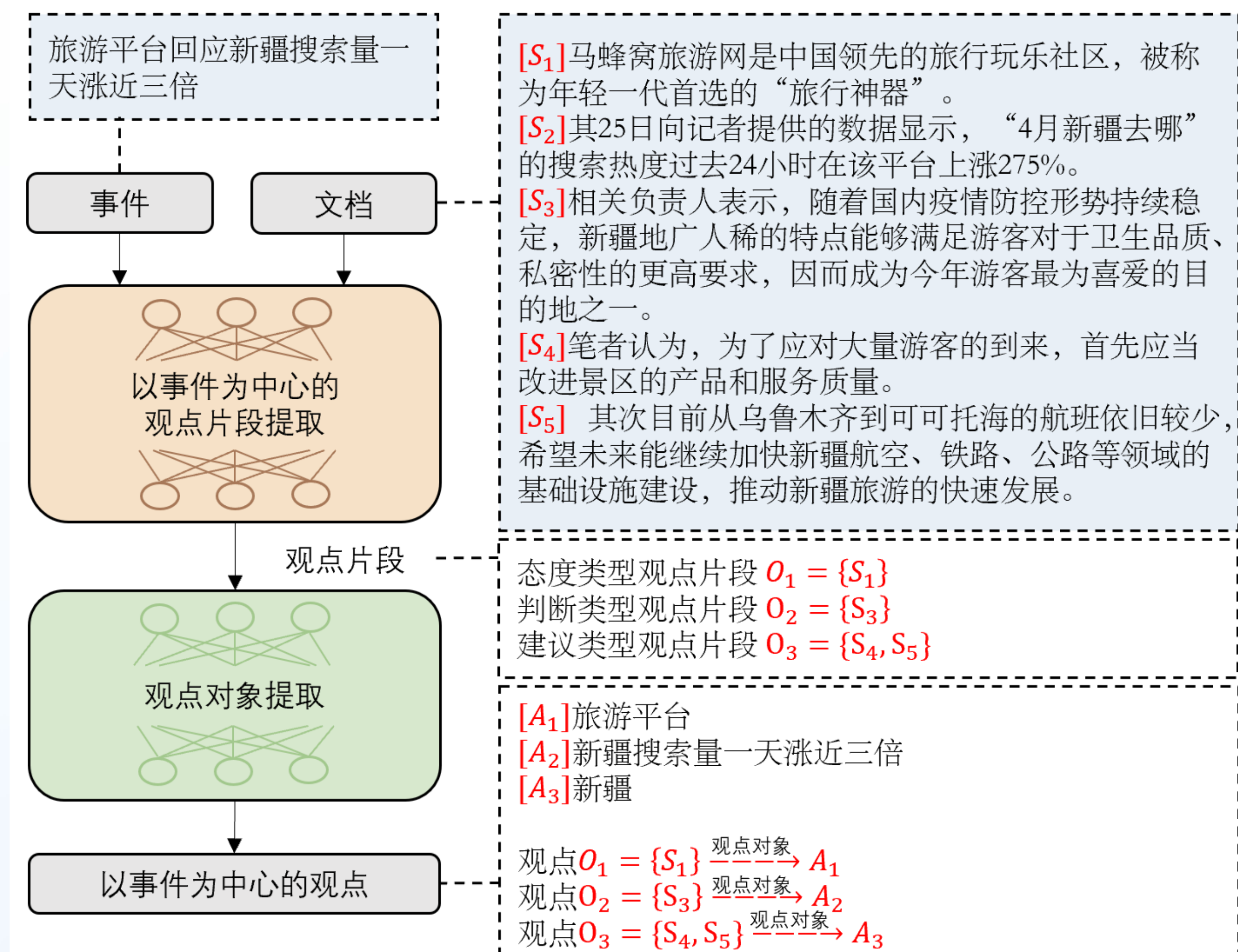
输入:

- 事件 $e = \{w_1, w_2, \dots, w_m\}$
- 文档 $d = \{s_1, s_2, \dots, s_n\}$

输出:

- 观点集 $T = \{..., (o_k, a_k), ... | e, d\}$
- 观点 $o = \{s_i, s_{i+1}, \dots, s_j | s \in d\}$
- 论元 $a = \{w_t, w_{t+1}, \dots, w_l | w \in e\}$

两阶段框架



ECO Bank

本文构建了以事件为中心的观点数据集, 包括中英双语。

Statistics on ECO Bank		Chinese	English
Document	Number	3000	2000
	Avg. Sents	15.2	20.3
	Avg. Opinion	2.6	4.5
Opinion	Number	7742	9058
	Ratio (%)	32.0	28.1
	Avg. Sents	1.9	1.3
Event	Number	821	167
	Avg. Tokens	6.4	7.7
Arguments	Events (%)	30.6	34.4
	Subevents (%)	24.9	11.7
	Entities (%)	44.5	53.9

实验

以事件为中心的观点挖掘是一个可行且具有挑战性的任务。

	ECO-ZH						ECO-EN					
	Segment level			Sentence level			Segment level			Sentence level		
	P	R	F ₁	P	R	F ₁	P	R	F ₁	P	R	F ₁
PairCls-SpanR	14.51	12.08	13.18	37.69	33.61	35.50	6.86	4.83	5.76	13.77	12.64	13.18
PairCls-MRC	13.45	11.19	12.22	48.99	43.67	46.13	14.67	10.42	12.19	33.12	29.32	31.10
Seq-SpanR	25.07	21.77	23.31	35.46	28.07	31.34	9.24	9.96	9.59	11.30	12.54	11.89
Seq-MRC	29.72	26.48	28.01	47.74	37.80	42.19	17.02	19.44	18.15	24.71	27.77	26.15
Human	86.96	86.02	86.49	79.46	94.23	86.22	72.59	82.10	80.83	86.78	86.07	86.42

结论

- 新任务: 以事件为中心的观点挖掘。
- 新数据集: ECO Bank。
- 基准: 两阶段框架。

联系方式

徐若曦 15022171058 ruoxi2021@iscas.ac.cn