Wei-ming LU, Jia-hui LIU, Wei XU, Peng WANG, Bao-gang WEI, 2020. EncyCatalogRec: catalog recommendation for encyclopedia article completion. *Frontiers of Information Technology & Electronic Engineering*, 21(3):436-447. https://doi.org/10.1631/FITEE.1800363

EncyCatalogRec: catalog recommendation for encyclopedia article completion

Key words: Catalog recommendation; Encyclopedia article completion; Product graph; Transductive learning

Corresponding author: Wei-ming LU
E-mail: luwm@zju.edu.cn
ORCID: https://orcid.org/0000-0002-0200-9215

Motivation

1. Online encyclopedias such as Wikipedia and Baidu Baike provide articles on many topics. However, there are still many entities that have not been authored. Although some articles have been edited several times, the content is still far from complete.

2. Providing various catalog items for articles is the key for encyclopedia article completion.

Main idea

1. Similar articles should have similar catalogs, so the catalog items of these articles can be considered as the candidate items for recommendation.

2. EncyCatalogRec is proposed to help generate a more comprehensive article in encyclopedias by recommending catalogs.

Method

1. We represent articles and catalog items as embedding vectors, and obtain similar articles via the locality sensitive hashing technology, where the items of these articles are considered as the candidate items.

2. A relation graph is built from the articles and the candidate items. This is further transformed into a product graph.

3. The recommendation problem is changed to a transductive learning problem in the product graph.

4. The recommended items are sorted by the learning-to-rank technology.

Major results

1. Results of the warm-start scenario for recommending different numbers of catalog items

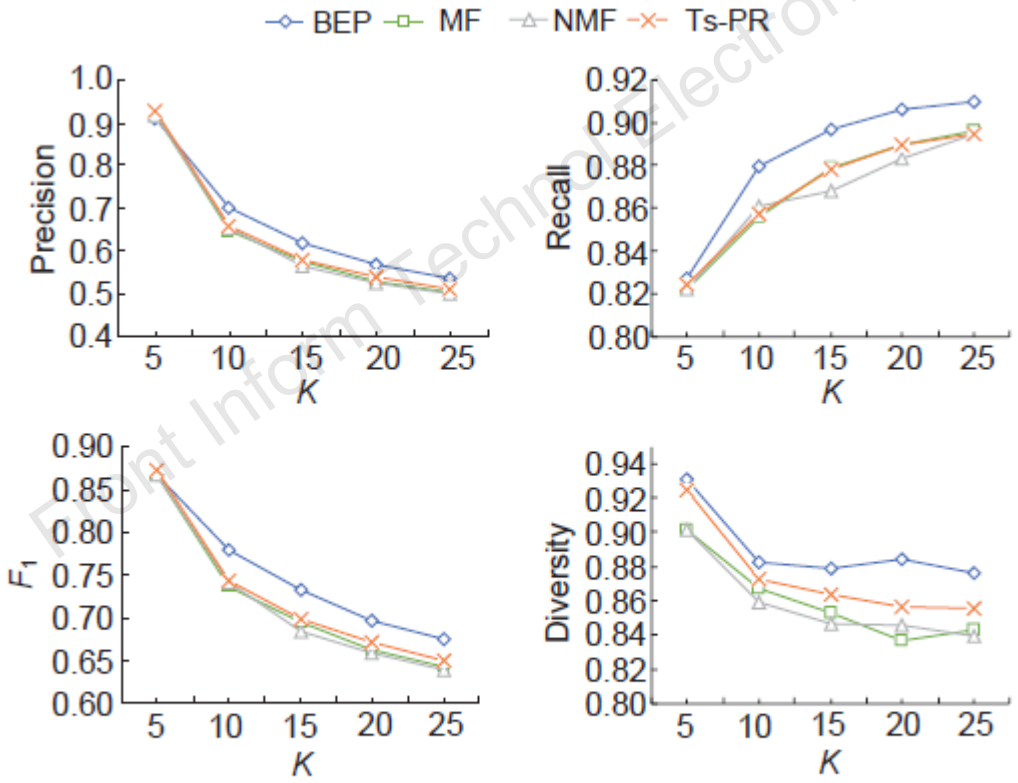

Fig. 5 Results of the warm-start scenario for recommending different numbers of catalog items

Major results (Cont'd)

2. Results of the cold-start scenario for recommending different numbers of catalog items

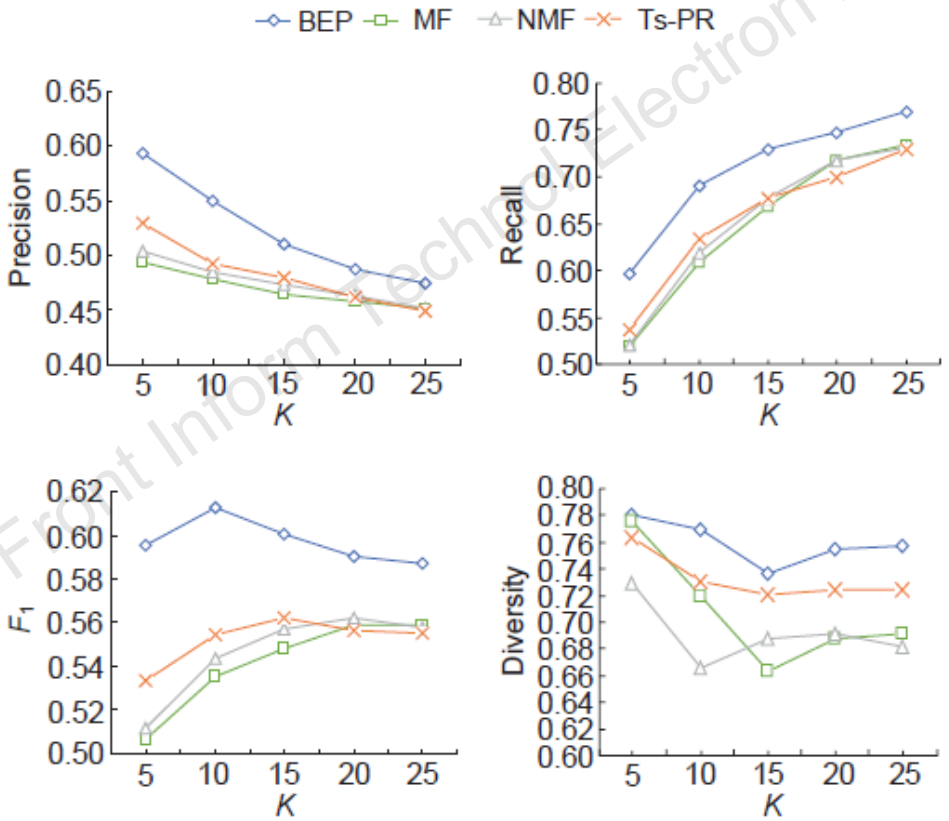

Fig. 6 Results of the cold-start scenario for recommending different numbers of catalog items

Major results (Cont'd)

3. Case studies

Table 3 An example of pericoronitis for catalog item recommendation

智齿冠周炎 (pericoronitis):	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 鉴别诊断 (antidiastole), 治疗 (treatment), 预后 (prognosis)
Method	**Recommended items**
MF#	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 诊断鉴别 (antidiastole), 诊断 (diagnose), 治疗 (treatment), 内容简介 (brief introduction of the content), 目录 (catalog), 版权信息 (copyright info.), 编辑推荐 (editor's choice)
NMF#	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 诊断鉴别 (antidiastole), 治疗 (treatment), 疾病治疗 (disease treatment), 版权信息 (copyright info.), 内容简介 (brief introduction of the content), 目录 (catalog), 编辑推荐 (editor's choice)
Ts-PR#	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 诊断鉴别 (antidiastole), 诊断 (diagnose), 疾病治疗 (disease treatment), 治疗 (treatment), 内容简介 (brief introduction of the content), 目录 (catalog), 版权信息 (copyright info.)
BEP#	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 治疗 (treatment), 诊断 (diagnose), 鉴别诊断 (antidiastole), 并发症 (complication), 疾病治疗 (disease treatment), 预防 (precaution), 预后 (prognosis)
MF△	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 诊断鉴别 (antidiastole), 疾病治疗 (disease treatment), 内容简介 (brief introduction of the content), 目录 (catalog), 版权信息 (copyright info.), 编辑推荐 (editor's choice), 作者简介 (about the author)
NMF△	病因 (cause), 临床表现 (clinical manifestation), 检查 (inspection), 诊断 (diagnose), 诊断鉴别 (antidiastole), 疾病治疗 (disease treatment), 内容简介 (brief introduction of the content), 目录 (catalog), 版权信息 (copyright info.), 作者简介 (about the author)
Ts-PR△	临床表现 (clinical manifestation), 检查 (inspection), 诊断 (diagnose), 诊断鉴别 (antidiastole), 治疗 (treatment), 疾病治疗 (disease treatment), 内容简介 (brief introduction of the content), 目录 (catalog), 版权信息 (copyright info.), 作者简介 (about the author)
BEP△	病因 (cause), 临床表现 (clinical manifestation), 治疗 (treatment), 诊断 (diagnose), 检查 (inspection), 并发症 (complication), 预防 (precaution), 预后 (prognosis), 内容简介 (brief introduction of the content), 目录 (catalog)

# denotes the warm-start scenario; △ denotes the cold-start scenario. Underlines denote the items selected for the warm-start scenario and wavy underlines denote the incorrectly recommended items in the results

Major results (Cont'd)

3. Case studies (Cont'd)

Table 4 An example of red maple for catalog item recommendation

红枫 red maple:	形态特征 (morphological characteristics), 生长习性 (growth habit), 分布范围 (distribution range), 主要价值 (chief value), 主要品种 (main varieties), 养护 (maintenance), 栽培技术 (cultivation technique), 红枫种植要点 (planting points for red maple), 植物文化 (plant culture)
Method	**Recommended items**
MF#	简介 (introduction), 形态特征 (morphological characteristics), 分布范围 (distribution range), 主要品种 (main varieties), 栽培技术 (cultivation technique), 植物文化 (plant culture), 信息 (information), 作品鉴赏 (works appreciation), 采集地 (collection sites), 内容简介 (brief introduction of the content)
NMF#	简介 (introduction), 形态特征 (morphological characteristics), 分布范围 (distribution range), 主要品种 (main varieties), 栽培技术 (cultivation technique), 植物文化 (plant culture), 采集地 (collection sites), 信息 (information), 地理分布 (geographical distribution), 生长习性 (growth habit)
Ts-PR#	简介 (introduction), 形态特征 (morphological characteristics), 分布范围 (distribution range), 主要品种 (main varieties), 栽培技术 (cultivation technique), 植物文化 (plant culture), 采集地 (collection sites), 信息 (information), 生长习性 (growth habit), 经济发展 (economic development)
BEP#	形态特征 (morphological characteristics), 分布范围 (distribution range), 主要品种 (main varieties), 栽培技术 (cultivation technique), 地理分布 (geographical distribution), 主要价值 (chief value), 生长习性 (growth habit), 分布情况 (distribution), 生长环境 (growth environment), 植物文化 (plant culture)
MF△	简介 (introduction), 形态特征 (morphological characteristics), 栽培技术 (cultivation technique), 采集地 (collection sites), 生长环境 (growth environment), 生长习性 (growth habit), 信息 (information), 作品原文 (content of work), 经济发展 (economic development), 自然资源 (natural resources)
NMF△	简介 (introduction), 形态特征 (morphological characteristics), 栽培技术 (cultivation technique), 采集地 (collection sites), 生长环境 (growth environment), 地理分布 (geographical distribution), 地理位置 (geographic location), 信息 (information), 作者简介 (about the author), 作品原文 (content of work)
Ts-PR△	简介 (introduction), 栽培技术 (cultivation technique), 采集地 (collection sites), 地理分布 (geographical distribution), 信息 (information), 生长习性 (growth habit), 主要价值 (chief value), 经济发展 (economic development), 作者简介 (about the author), 自然资源 (natural resources)
BEP△	形态特征 (morphological characteristics), 地理分布 (geographical distribution), 栽培技术 (cultivation technique), 采集地 (collection sites), 生长习性 (growth habit), 主要价值 (chief value), 生长环境 (growth environment), 植物文化 (plant culture), 分布情况 (distribution), 信息 (information)

# denotes the warm-start scenario; △ denotes the cold-start scenario. Underlines denote the items selected for the warm-start scenario and wavy underlines denote the incorrectly recommended items in the results

Conclusions

1. EncyCatalogRec was proposed to help generate a more comprehensive article by recommending catalogs. This is a good starting point for authors to complete articles in encyclopedias.

2. We compared our approach with other methods in both warm- and cold-start scenarios with a dataset created from Baidu Baike. The results proved that EncyCatalogRec can achieve the state-of-the-art performance, and can recommend useful catalog items for encyclopedia article completion.