Medical Entity Disambiguation with Medical Mention Relation and Fine-grained Entity Knowledge

Wenpeng Lu[1,4], Guobiao Zhang[1,4], Xueping Peng[2], Hongjiao Guan[1,4,*], Shoujin Wang[3]
[1]Key Laboratory of Computing Power Network and Information Security, Ministry of Education,
Shandong Computer Science Center (National Supercomputer Center in Jinan)
Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2]Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, Australia
[3]Data Science Institute, University of Technology Sydney, Sydney, Australia
[4]Shandong Provincial Key Laboratory of Computer Networks,
Shandong Fundamental Research Center for Computer Science, Jinan, China
{wenpeng.lu,hongjiao.guan}@qlu.edu.cn, zgb_stubbron@163.com
{xueping.peng,shoujin.wang}@uts.edu.au

Guobiao Zhang

LREC-COLING 2024

MED (Medical Entity Disambiguation)

- Insulinum is an important factor in the treatment of DM.

Medical Entity Disambiguation (MED) refers to the task of mapping medical mentions in medical text documents to their corresponding entities in a knowledge base (KB).

Challenge 1:
Neglect more knowledge information in KBs

Existing medical entity disambiguation methods primarily rely on a single type of knowledge and do not effectively integrate the knowledge embedded within medical KBs, which inadequately models the semantic representation of medical mentions and the semantic representation of candidate entities.

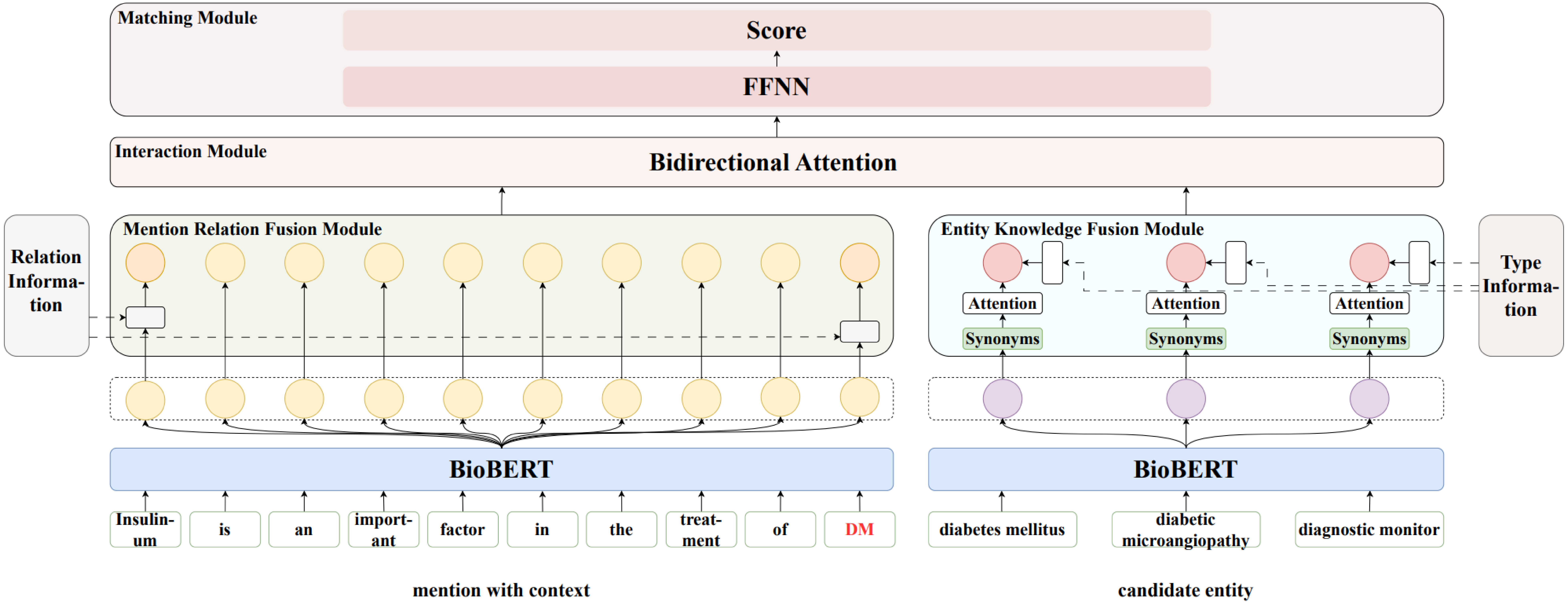
Challenge 2:
Overlook essential interactions

Existing medical entity disambiguation methods overlook essential interactions between medical mentions and candidate entities.

Contribution

We propose a novel approach for medical entity disambiguation with medical mention relation and fine-grained entity knowledge (MMR-FEK). In contrast to existing works, our approach mines more additional relation knowledge between mentions and more fine-grained entity knowledge to enhance their representations, and captures more interactions between mentions and entities.

Model


Matching Module
Score
FFNN
Interaction Module
Bidirectional Attention
Relation Information
Mention Relation Fusion Module
Entity Knowledge Fusion Module
Type Information
Attention
Attention
Attention
Synonyms
Synonyms
Synonyms
BioBERT
BioBERT
Insulinum
is
an
important
factor
in
the
treatment
of
DM
diabetes mellitus
diabetic microangiopathy
diagnostic monitor
mention with context
candidate entity

Datasets

MedMentions is the most popular and largest biomedical entity disambiguation dataset, which consists of 4,392 abstracts from PubMed, with over 350,000 mentions linked to UMLS concepts.
BC5CDR consists of 1,500 articles from PubMed, with 4,409 annotated chemicals and 5,818 diseases. It contains over 28,000 mentions linked to MeSH concept, which are mapped to UMLS ones by us.

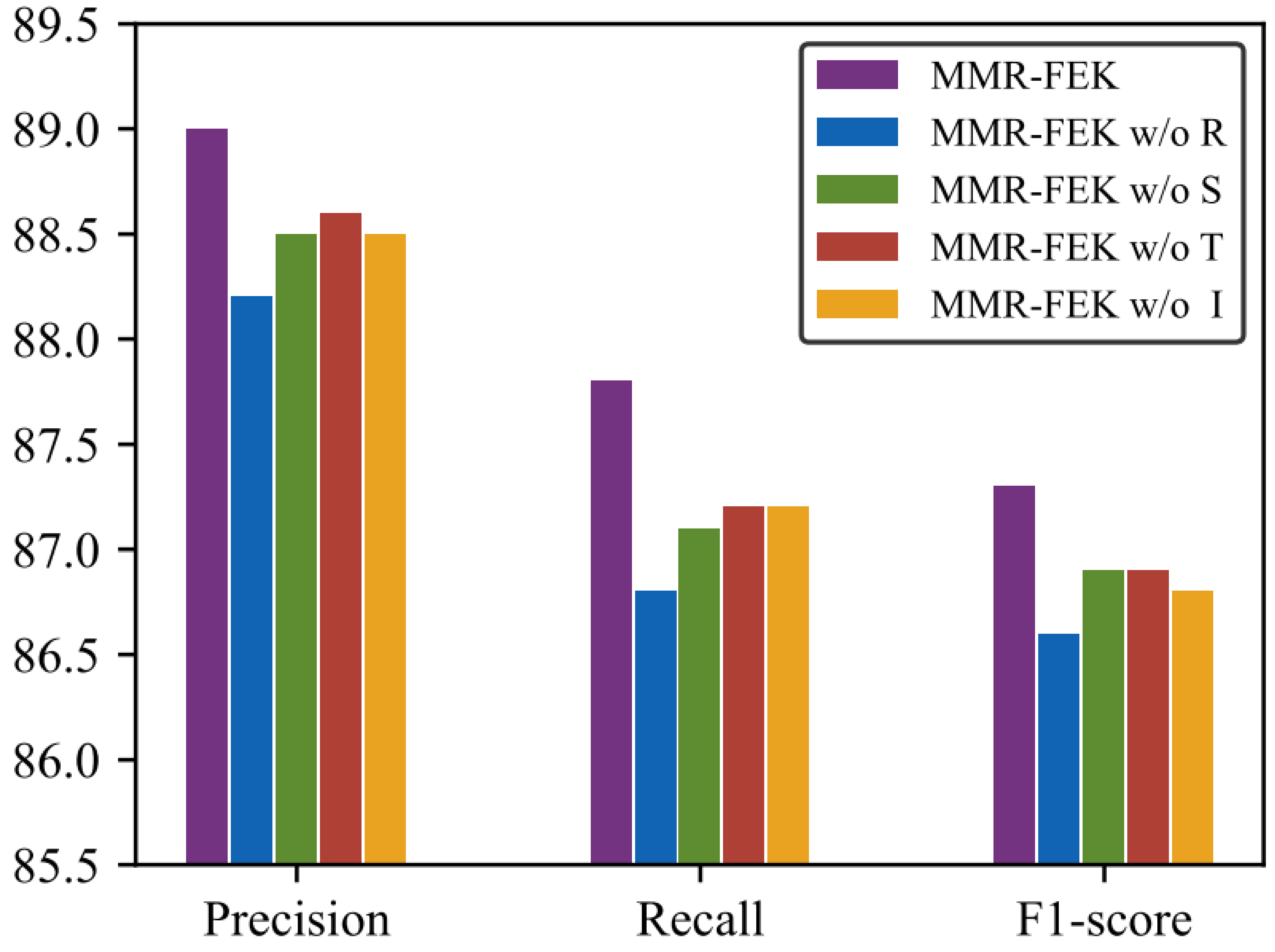
Dataset	Statistics	Train	Dev	Test
MedMentions	#Documents	2,635	878	879
	#Mentions	211,029	71,062	70,405
	#Entities	20,830	6,941	6,953
BC5CDR	#Documents	900	300	300
	#Mentions	17,135	5,710	5,714
	#Entities	5,489	1,830	1,830

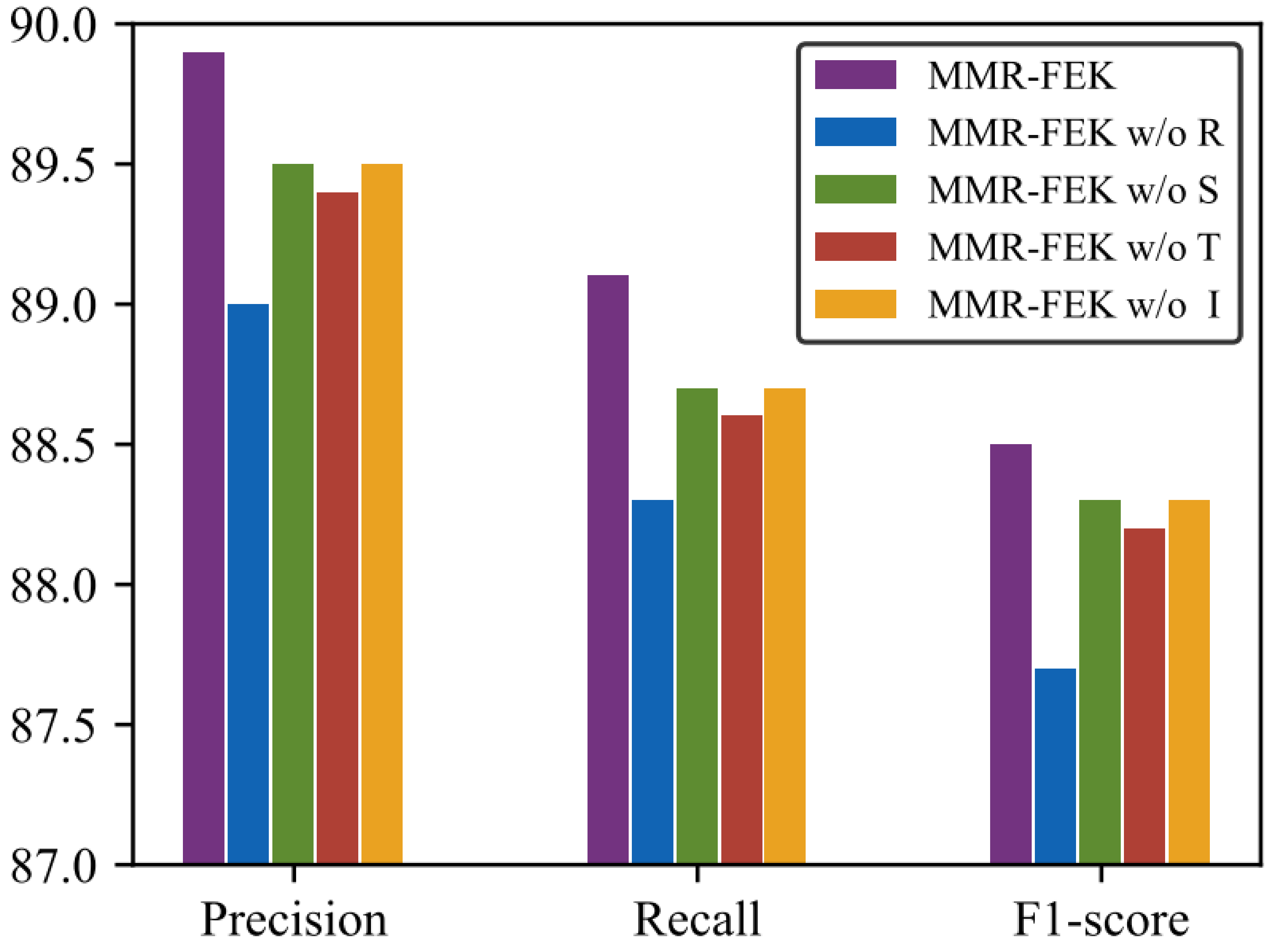
Table 1: Statistics of datasets.

Experiment Result

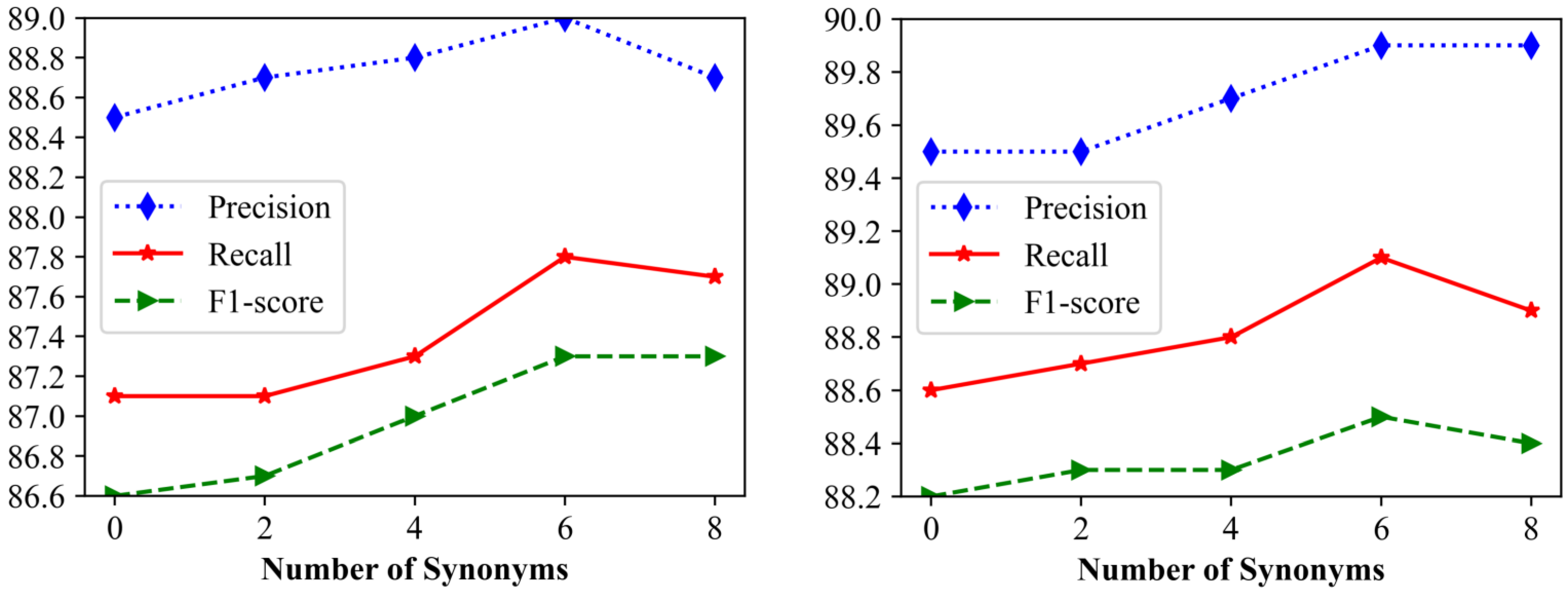
Model	MedMentions			BC5CDR		
	Precision	Recall	F1	Precision	Recall	F1
NCEL	OOM	OOM	OOM	65.7	67.3	65.2
BIOSYN	OOM	OOM	OOM	69.1	68.8	70.1
SAPBERT	53.1	56.4	52.3	70.2	72.9	72.4
Dual Encoder	62.9	67.4	65.6	84.8	82.1	83.0
Zhu	66.5	68.1	65.4	83.2	84.0	81.1
LATTE	88.2	86.5	85.6	88.2	87.0	86.3
B-LBConA	88.5	87.1	86.5	89.1	88.2	87.3
MMR-FEK (our)	**89.0***	**87.8***	**87.3***	**89.9***	**89.1***	**88.5***
Improvement(%)	0.5	0.7	0.8	0.8	0.9	1.2

Ablation Study





Influence of Number of Synonyms



Code

The code is available from my GitHub page:

https://github.com/Stubborn-z/MMR-FEK

Thanks