

第3章 因果关系识别

大数据环境下,大量的社交媒介以爆炸式的方式生产着代表人们日常语言和行文的文本数据。这些语言文本中蕴藏着大量的人们对生活经验的总结和事物间关联关系的表述。如何从海量数据中基于以往经验数据提取事物间的联系,基于事物间得相关关系提取文本数据中隐含的因果关系是本节讨论的重点。

3.1 显式因果关系

3.1.1 因果关系表示形式

1. 因果关系的语义形式

从汉语言学角度,因果关系是文本中概念或词语间存在的原因和结果关系,是经典的语义关系,以因果句的形式表述出来。汉语言文本中因果句主要包含三中语义形式^{[22][23],[24]}:

① 事理关系:是客观事物关系的映射,由事物的相关性和因果性组成。事物的相关性具有并存关系;因果性具有主动造成与被动引起的关系。

② 认知关系:指基于语言形式对客观事物间关系的一种理性认识。一般表示为由关联词构成的复句形式,即反应事理关系由反应认知关系。

③ 心理关系:属于世界观、人生观范畴,是人类认知的一种主观态度,心理反应,属于语用意义。

文本数据中因果关系是客观事物间存在的一种关系,条件性、特定性、时序列性为其特点。在文本数据中从语义形式到具体的事物间因果关系的映射,即为因果关系的抽取。

2. 因果关系的表达形式

根据汉语言表达习惯,在汉语言表达中基于短语、单句、复句、段落等语言单位都可以表达事物间的因果关系。综观国内外学者的研究,从表达单位上可以将因果关系分为如下几种:

① 句内因果关系:在一个句子中由语义之间的关系表达事物间因果关系的语句。单个句子内存在的因果关系包括简单因果和复杂因果。如:“浙江大面积降雪导致人民深受其害!”。

② 句间因果关系：文本中表达的因果关系存在与多个句子之间，需要分析多条语句才能确定事物间的因果关系。如：“全国大面积降温，带来了冻雨现象。……浙江民众受到财产损失！”

③ 段落间因果关系：事物间的因果关系隐含在段落之间，原因部分和结果部分可能跨段落存在，需要根据上下文语义识别此种因果关系。

④ 篇章间因果关系：文本中事物间的因果关系跨篇章存在，需要结合多篇文本结合上下文语义识别的因果关系。

基于文本中因果关系的语意表达类型的不同，文献^[25]将汉语文本中的含有因果关系的语句归纳为以下三种表达形式：

① 无标识的因果句和有标识的因果句。将语句中含有因果提示词且能表征语句为因果关系的语句称为“有标识因果关系句”；反之，语句中未有因果标记提示词的语句称为“无标记因果句”。

② 明确因果句和模糊因果句。对于文本中含有因果提示词的语句，如果总能表征语句含有因果关系则称为明确性因果句；若文本中含有因果标记提示词，但不能确定此语句为因果关系，则为模糊因果句。

③ 隐式因果句和显式因果句。显式因果句为可以直接从字面上通过因果提示词识别其因果实体；然而不能通过特定词、句法分析、模式分析等技术，只能通过相关知识结合统计分析和上下文语义才能识别出因果关系的语句为隐式因果句。

3.1.2 汉语中的因果关系词

英语的语义注重形式上表达，而汉语侧重于语句中语义的阐述，再加上国内对文本中因果关系的研究起步较晚，以及因果关系的句法模式和语言模式归纳不完整造成国内因果关系提取的准确率和质量都略显偏低。本文结合汉语言文学知识，将汉语言语句中能表征因果句式的关系词称为“因果提示词”。中文中能表中因果关系的词语主要包括：动词、介词、连词和副词。通过分析知网中义原的属性，归纳出表因果的义原、表结果的义原和表状态的义原都可以表达概念间的因果关系。故本文基于知网义原属性归纳的动词因果提示词概括如下：

表 3.1 动词

义原属性		动词
Cause_Result	Result_In {无取值约束}	导向、勾起、导致、产生、造就、造成、挟带、酿成、引发、引来、诱致、引致、滋生、诱发、招致、决定、致使、归于、带来、令人、触发、引出、引起、
	Result_In {结果有取值约束}	召祸、惹祸、致残、致死、成灾、成瘾、致病、成癖
	Influence	推进、波及、关系、牵动、浸染、渗透、渗入、攸关、影响、推动、作用
	Result_From	出于、取决、出自、发源于、缘于、发自、源于、在于、来自、来源于、起源于
	Base_on	立足、根源于、根植于
Make_Act	Cause To Do	使、促发、使得、促进、诱杀、促使、指引、促成、引入、引导、诱使、诱导、诱惑
	Cause To Do {结果有取值约束}	普及、放活、发汗、固定、放飞、费事、解密、架空、
State	State_Mental	受伤、喜欢、后悔、埋怨、生气
Means		毒死、杀死、气死、饿死、风干

结合汉语言文学知识,借鉴许义胜^[27]、邢福义^[28]、倪重阳^[29]、彭湃^[30]等对汉语中因果关系词的研究成果,本文归纳出词性为:连词、副词和介词等的因果提示词如表 3.2 所示:

表 3.2 副词、连词和介词

词性	模式	提示词
conj	Single_conj	因(为)、于是、由于、从而、既(然)、因而、以致(于)、如果、所以、只要、因此、致使、故、以至(于)
	Math_conj	{因为,为此}、{既(然),所以}、{由于,为此}、{由于,以至(于)}、{既(然),却}、{由于,从而}、{既(然),就}、{由于,故}、{如果,就}、{既(然),因此}、{由于,于是}、{只要,就}、{由于,致使}、{因为,所以}、{因为,因此}、{由于,所以}、{因为,以致(于)}、{由于,以致(于)}、{因为,因而}、{(之)所以,(是)因为}、{因为,于是}、{只有 除非,才}、{因为,致使}、{如果,那么/则}、{因为,故}、{由于,因此}、{因为,故而}、{因(为),以至(于)}、{因为,从而}、{由于,因而}、{(之)所以,(是)由于}
prep		为了、因(为)、为、按照、以、靠、凭、比、依据、由于
adv		为此、以便、才、以免、以免

3.1.3 因果关系表达结构分析

基于因果关系提示词从文本中识别因果关系就需要归纳总结汉语言的语法结构与表达习惯。结合汉语言文学知识，概括出汉语言文本中因果关系句的共性基本语法结构，如表 3.3 所示：

表 3.3 因果关系表达结构

表达结构	Causes → Effects
一因一果	形如：“C→E”结构 例句：“因为很类，所以我想哭”。
多因多果	形如：“{C ₁ , C ₂ ... C _m } → {E ₁ , E ₂ ... E _n }”结构 例句：“由于气温骤降，多个地区连续多日的暴雨，造成洪水泛滥，居民财产受损”。
一因多果	形如：“C → {E ₁ , E ₂ ... E _n }”结构 例句：“由于洪水泛滥，导致多地区受灾，民众财产受损严重”。
多因一果	形如：“{C ₁ , C ₂ ... C _m } → E”结构 例句：“因为市场不景气，导致工厂倒闭，民工失业等现象”。

3.2 文本中因果关系模式归纳

国内，基于汉语文本实体关系的提取大多依赖于文本的语法结构，构建相应的模式，然后基于模式匹配方法在特定领域进行实体关系的识别。可见模式匹配方法是现阶段在自然语言处理中占据重要地位。对文本数据中因果关系的提取主要涉及因果句的判断和因果关系实体对的抽取。通过对汉语文本因果句进行系统、全面的研究和分析，本文基于因果关系提示词进行因果句的识别，并依据汉语语句的语法结构，总结和归纳出各类提示词相应的因果关系句法模式。

3.2.1 连词对应的句法模式

连词是用来承接词、句子或者段落，表示某种逻辑语义的虚词，可用来表示连接成分间的因果、并列、选择、假设、承接、比较、转折和让步等关系^[31]。本文主要分析表征因果关系的部分，如：“因为、因此、由于、所以、是故、以便、以致”。基于因果连词分析其对应的句法模式。

1、配套连词

配套连词是两个连词一起使用，共同表征概念间的因果关系。根据连词间搭配模式的不同，将其分为由因到果型和由果溯因型：

(1) 由因到果

根据汉语言文学知识，此类连词的第一个提示词后面为原因部分，第二个提

示词后面为结果部分。相应的因果连词有：{〈因为, 所以〉、〈因为, 从而〉、〈因为, 因此〉、〈只有|除非, 才〉、〈因为, 以致[于]〉、〈如果, 那么|则〉、〈因为, 因而〉、〈既[然], 所以〉、〈因为, 于是〉、〈既[然], 因此〉、〈因为, 致使〉、〈由于, 以至[于]〉、〈因为, 故〉、〈由于, 从而〉、〈因为, 故而〉、〈因(为), 以至(于)〉、〈由于, 故〉、〈由于, 因此〉、〈由于, 于是〉、〈由于, 所以〉、〈由于, 致使〉、〈由于, 以致〉、〈由于, 因而〉}。其相应句法模式表示如下：

$$\langle \text{Conj}_k \rangle \{ \text{Cause} \} , \langle \text{Conj}_k \rangle \{ \text{Effect} \}$$

其中, $\langle \text{Conj}_k \rangle$ 代表因果提示词, 如: $\langle \text{Conj}_k, \text{Conj}_k \rangle = \{ \text{因为, 所以} \}$, $\{ \text{Cause} \}$ 表示原因部分, $\{ \text{Effect} \}$ 表示结果部分。例句: “〈因为〉{爱}, 〈所以〉{爱}”。

(2) 由果溯因

此类因果联系表征的因果句, 结果部分紧跟在第一个因果连词后, 原因部分紧跟在第二个因果连词后。中文中相应的因果连词有: {〈[之]所以, 因为〉, 〈[之]所以, 缘于〉, 〈[之]所以, 由于〉}。其相应的句法模式表示如下:

$$\langle \text{Conj}_k \rangle \{ \text{Effect} \} , \langle \text{Conj}_k \rangle \{ \text{Cause} \}$$

例: $\langle \text{Conj}_k, \text{Conj}_k \rangle = \{ \text{之所以, 是因为} \}$ 。例句: “我〈之所以〉{这么用功的学习}, 〈是因为〉{学习能够改变命运}”。

2、单连词

综合分析得知, 由单连词表达的因果句其因果提示词(单连词)成分分布在语句的前分句或者语句中间。本文对其进行分别讨论。

① 单连词位于前分句

归纳汉语言文学对单连词的研究, 由{因(为)、如果、由于、只要、既(然)}等表征的因果句, 因果提示词后的分句, 即为原因部分。其对应的句法模式表示如下:

$$\langle \text{Conj} \rangle \{ \text{Cause} \} , \{ \text{Effect} \}$$

此种句法模式的提取需要依据句法依赖关系并结合上下文语义进行因、果部分边界的识别。例句: “〈只要〉{还有一线希望}, {我们就不能放弃}”。

② 单连词位于分句中间

在自然语言表达中，此种连词位于语句的中间位置，连接原因部分和结果部分。根据其引导的成分不同，将其分为由果溯因和由因到果两种模式。

(1) 由因到果

由此种因果连词表达的因果句，原因部分在提示词之前，提示词之后为结果部分。相应的因果连词有：{因此、所以、致使、以致[于]、故、因而、于是、从而、以至[于]}。对应的句法模式表示如下：

$$\{\text{Cause}\}, \langle \text{Conj} \rangle \{\text{Effect}\}$$

例句：“小明{睡过头}了，〈以至于〉{没买着回家的票}”。

(2) 由果溯因

汉语言中由 {因为、由于} 等因果连词表达的因果句（因果连词位于分句中间），其提示词前的分句表结果部分，其后部分表造成前述结果的原因。句法模式表示为：

$$\{\text{Effect}\}, \langle \text{Conj} \rangle \{\text{Cause}\}$$

例如：“大雪{挡不住人们回乡的脚步}，〈因为〉{回家是一种信仰!}”。

3.2.2 动词对应的句法模式

动词 (Verb) 能够表示事物或人动作或状态的一种动态变化。文献^{[32], [33]}指出动词隐含因果性特征，表示人或事物一个状态的改变导致另一状态的改变。本文通过对知网中义原间属性的分析，并结合 Nedjalkov 的研究成果将汉语中的因果动词分为：表示结果的因果动词、提示性因果动词、表结果产生式的因果动词^[34]。

1、提示性因果动词

此类动词自身不表征任何因果成分，在因果句中仅作为具有因果语义的动词性因果提示词。根据其在汉语言表达习惯中，由此种动词表达的因果句可分为：由果溯因式和由因到果式。由此类提示词表达的因果句一般为明确因果句。其表示结构如表 3.4 所示：

表 3.4 提示性因果动词句法模式分析

由因到果	
因果成分	{原因部分}, <动词性因果关系词>{结果部分}
提示词	{导向、予以、导致、使动、勾起、使得、产生、指引、造成、诱导、造就、酿成、引入、引发、促成、引起、促使、引来、促发、引致、推进、诱发、推动、诱致、攸关、招致、促进、致使、诱惑、滋生、引导、归于、作用于、决定、渗透、牵动、令人、影响、引出、浸染、带来、使、波及、挟带、关系、触发、渗入、诱使}
句法模式	{Cause} , <Verb> {Effect}
例句:	{潮湿的环境}容易<滋生>{细菌}。
由果溯因	
因果成分	{结果部分}, <动词性因果关系词>{原因部分}
提示词	根源、出于、缘于、发源于、发自、源于、取决、来源、起源、来自、取决于、立足
句法模式	{Effect} , <Verb> {Cause}
例句	一个人的{成功}<取决于>后天的{努力}。

2、充当结果的动词

充当结果的动词自身可以表征因果关系中的结果部分,此种因果句原因部分有时缺失。在知网中主要有以下词语组成:喜欢、埋怨、普及、后悔、固定、生气、解密、受伤、费事、召祸、放飞、致病、发汗、致残、惹祸、成癖、致死、成灾、放活、成瘾、架空。汉语中相对应的句法模式:

$$\{...\} <Verb>_{Effect}$$

其中, {...} 表示不确定性成分,也可以表示没有原因部分, <Verb>_{Effect} 表示此“词组”既是因果提示词,有充当因果句中的结果部分。此种因果提示词表示的因果句对于因果成分边界的界定具有极大的模糊性,现有的国内外研究对此种因果句中因果关系的提取技术还有待提高。本文提出基于统计学和概率学知识对其进行因果关系实体模式对的提取。

3、表结果产生式的动词

知网中义原属性表示语义间概念的影响方式,如:“杀死、饿死、风干、毒

死”等，在汉语言学中表征因果关系的影响方式。此种动词自身就阐述了原因部分作用于结果部分产生的后果。比如：“根据对案发现场的探察，受害者被毒死的”。相应的句法模式表示如下：

$$\{Cause\} \langle Verb \rangle \{Effect\}$$

根据句法模式对例句进行解释，结果为： $\{毒\} \rightarrow \{死\}$ 。根据语言学知识，可知此种动词表征的因果句属于显式模糊因果关系，对于因果各部分的边界界定，以及原因部分和结果部分的成分识别提出了极大的挑战。

3.2.3 介词对应的句法模式

介词 (Prep) 又称作前置词，表示词与句、词与词之间关系的虚词，介词可以用来表地点、表示时间、表原因、表示数量、表示方式。介词作为因果句的提示词其引导的句法模式为： $\langle 介词 \rangle \{原因部分\} \{结果部分\}$ 。其对应的句法模式表示为：

$$\langle Prep_k \rangle \{Cause\}, \{Effect\}$$

其中， $\langle Prep_k \rangle = \{为、按、为了、按照、依据、凭借、依靠、因(为)、比、由于\}$ 。例句：“ $\langle 凭借 \rangle$ 其过人的天赋，他在此次比赛中拿到了冠军奖杯”。由原因介词表达的因果句，属于显式因果关系中的模糊因果关系，因为汉语中介词可以表征多种语义，只能根据上下文语义进行判断改语句是否真正表征因果关系。本文提出了基于朴素贝叶斯进行分类，有效的提高了基于介词因果关系抽取的准确率。

3.2.4 副词对应的句法模式

副词 (Adverb) 指在句子中表征概念行为或状态特征的词，其中有表征关系的副词。本文主要研究其属于因果关系的关系副词（为此、以免、才、以便）。其对应的句法模式表示如下：

$$\{Cause\}, \langle Adverb \rangle \{Effect\}$$

同样由于其词性的多义性，由副词表征的因果句也属于显式模糊因果关系。关系副词作为因果标识词有时也和因果连词搭配使用，表示概念间的因果关系。如： $\{ \langle 因为, 为此 \rangle, \langle 只要, 就 \rangle, \langle 由于, 为此 \rangle, \langle 如果, 就 \rangle, \langle 既[然],$

就)、〈既[然], 却〉}。其对应的句法模式为:

$$\langle \text{Conj} \rangle \{ \text{Cause} \}, \langle \text{Adverb} \rangle \{ \text{Effect} \}$$

此种模式属于配套因果连词的一部分, 为显式明确因果关系。例如: “信息化时代, 〈因为〉信息的可知度越高{产品的销路越广}, 〈为此〉所有的企业都要搭乘{互联网+}这辆顺风车”。

3.2.5 明确因果与模糊因果分类

综上对连词、动词、副词和介词表征因果语义的分析, 连词可显式明确的表征语句中概念的因果关系, 并能通过语法结构进行因、果各部分的界定与抽取。而副词、动词、介词由于其词性的多义性, 作为因果提示词表征因果句时存在歧义性, 因此由介词、动词、副词表征的因果句属于显式模糊因果句。为更方便的区分因果标识词与句法模式在明确因果句或模糊因果句中的对应关系。本文做如下细分:

表 3.5 模糊因果句法模式

词性	提示词	句法模式
Prep	为了、依据、为、按照、因[为]、按、依赖、照、比、凭借、由于	由因到果居端式 $\langle \text{Prep} \dots \rangle \{ \text{Cause} \}, \{ \text{Effect} \}$
Verb	牵动、导向、使动、导致、勾起、引入、指引、使、予以、产生、促成、造成、引导、造就、促使、酿成、引发、渗透、促进、引起、诱导、引来、促发、引致、诱发、推进、诱致、推动、招致、影响、致使、滋生、归于、作用、使得、决定、攸关、令人、引出、浸染、带来、挟带、触发、关系、渗入、诱惑、波及、诱使	由因到果居中式 $\{ \text{Cause} \}, \langle \text{Verb} \text{Adverb} \dots \rangle \rightarrow \{ \text{Effect} \}$
Adverb	以免、以便、为此、才	
Verb	根源于、取决、来源于、出于、取决于、缘于、在于、出自、起源于、来自、发源于、发源于、根源于、立足[于]	由果溯因居中式 $\{ \text{Effect} \} \langle \text{Prep} \dots \rangle \{ \text{Cause} \}$

表 3.6 明确因果句法模式分析

词性	提示词	句法模式
S_Conj	既[然]、因[为]、如果、由于、只要	由因到果居端式 <Conj...>{Cause}, {Effect}
	于是、所以、故、致使、以致[于]、因此、以至[于]、从而、因而	由因到果居中式 {Cause}, <Conj...>{Effect}
	因为、由于	由果溯因居中式 {Effect}<Conj...>{Cause}
M_Conj	〈因为, 从而〉、〈因为, 为此〉、〈既[然], 所以〉、〈因为, 为此〉、〈由于, 为此〉、〈只有 除非, 才〉、〈由于, 以至[于]〉、〈既[然], 却〉、〈如果, 那么 则〉、〈由于, 从而〉、〈既[然], 就〉、〈既[然], 因此〉、〈如果, 就〉、〈只要, 就〉〈因为, 所以〉、〈由于, 于是〉、〈因为, 因此〉、〈由于, 故〉、〈因为, 以致[于]〉、〈因为, 因而〉、〈由于, 因此〉、〈因为, 于是〉、〈由于, 致使〉、〈因为, 致使〉、〈由于, 以致[于]〉〈因为, 故〉、〈因[为], 以至[于]〉、〈由于, 所以〉、〈因为, 故而〉、〈由于, 因而〉	由因到果配套式 <Conj>{Cause}, →<Conj>{Effect}
	〈[之]所以, 因为〉、〈[之]所以, 由于〉、〈[之]所以, 缘于〉	由果溯因配套式 <Conj>{Effect}, →<Conj>{Cause}

3.3 隐式因果关系

隐式因果关系是指：在数据语料中没有显式因果提示词，而隐含因果关系的语句。汉语言文学中，可以将文本中包含的因果关系概括为两种形式：

① 形合法：汉语言语句中具有显式的形式标记（因果提示词），显式地表达概念间的因果关系；

② 意合法：汉语言语句中没有表征概念间因果关系的提示词，含蓄地表达概念间的因果关系。

隐式因果关系识别一般需要基于概念间上下文语义进行推断、理解，采用相关知识判断两概念间是否存在因果关系。隐式因果关系在汉语语句中无显式的词法、语法特征、缺乏统一的模式，在加上中文语义的歧义性、多义性，导致隐式因果关系识别效果不够理想。

隐式因果关系实质还是两概念间的相关关系。本文提出从概念间相关关系着

手,然后将概念间因果关系发生的条件作为判断依据,从而达到隐式因果关系的识别。汉语言中因果关系在哲学、逻辑学和语言学的研究中理论成果已相当成熟,本文从汉语言文学、哲学、统计学的角度分析事物因果关系发生的条件,主要包括以下几个方面:

① 时序性:在时间上,原因事件早于结果事件。Hume^[35]提出因果关系是两个事件间的固有联系,在时间和空间上是相邻的,原因时间导致了结果事件的变化,所以原因事件优先发生。

② 依赖性:结果事件由原因事件所造成,即结果事件的状态受原因事件影响。在逻辑学和哲学的因果关系操纵理论中指出:通过对原因事件的控制可影响的发生程度^[36]。

③ 提升性:原因的发生一定程度上增加了结果产生的概率。Suppe指出:如果事件 A 发生提高了 B 产生的概率,如: $P(B|A) > P(B)$, 那么事件 A 即为事件 B 发生的原因。由于忽略了事件间发生的时间先后性以及事件本身所固有的相关性,因此该定义存在一定的歧义性。

近年来国内外,基于文本数据对隐式因果关系已做了大量研究,但是效果都不甚理想。再加上中文语义的多义性、歧义性、模糊性,给基于中文文本的隐式因果关系挖掘提出了巨大挑战。基于开放领域做因果关系的判别,所以主要研究显式因果关系的提取与识别。

3.4 本章小结

本章主要基于哲学、汉语言文学、逻辑学等讨论在汉语言中存在的因果关系。基于汉语言文学知识总结并归纳出汉语言中存在的因果关系提示词,并分别进行讨论;然后总结汉语中基于因果提示词表达的因果关系的句法结构,并归纳出相应的句法模式;对隐式因果关系进行了简单分析。