

# 基于CFC(正确性信心指数)的学习型可信赖机器翻译系统

李应潭

中国科学院 TMT 研发组

Email:lyt@sia.ac.cn

**摘要** 本文阐述基于正确性信心指数(CFC)的学习型可信赖机器翻译系统(TMTSoLA)的基本理论问题——CFC在TMTSoLA里的地位和作用问题。讨论了二分数制和百分数制这两种CFC策略,指出CFC与传统翻译系统的翻译正确率(CF)存在本质区别,但是对于施行百分制CFC策略的TMT系统而言,它的CFC在计算方法上却与CF有相似之处,但施行百分制CFC策略的TMT系统与非可信赖机器翻译系统的操作仍有本质的区别。CFC是针对各个层次的具体语言对象计算的,而CF却没有具体语言对象针对性。为了能够计算出针对各个层次具体语言对象实际翻译算法的CFC,TMT系统必须具有具体的词语、句型和数据驱动的翻译模型,而不是笼统的翻译模型,否则CFC将失去意义。最后还论述了CFC与内嵌学习方法的本质联系。

**关键词** 可信赖机器翻译,正确性信心指数(CFC),基于CFC的可信赖机器翻译,学习能力

## Confidence Factor of Correctness Based Trustworthy Machine Translation System of Learning Ability

Li Yingtan

TMT R&D Group,CAS

Email:lyt@sia.ac.cn

**ABSTRACT** In this paper author presents a Confidence Factor of Correctness based Trustworthy Machine Translation System of Learning Ability(short as CFC based TMTSoLA).We talk about its basic theoretical problems—the position and function of CFC in TMT System. At first we define the CFC as a measure of the confidence to a translation of certain sentence, phrase or context for a MT system. CFC is different from the CF(Correctness Factor), later is a measure for the whole translation result of a MT system. There are two numeric system for CFC—the binary system and percentage system. We discuss their relation and distinguishes between binary systemic CFC and percentage systemic CFC. But there is no relation between CF and CFC

**KEYWORD** TMT(trustworthy machine translation),CFC(confidence factor of correctness),CFC based TMT, Learning Ability, TMTSoLA (Trustworthy Machine Translation System of Learning Ability)

## 1 引言

CFC based TMTSoLA (Confidence Factor of Correctness based Trustworthy Machine Translation System of Learning Ability)——基于 CFC 的有学习能力的可信赖机器翻译系统，是 TMT 系统的一个工程实践形式。

TMT 是我们于上一世纪末基于可信赖人工智能 (TAI) 原理<sup>[1]</sup>和可信赖机器翻译 TMT 的基础之上提出来的。

首先，可信赖人工智能理论证明：为了实现可信赖性这个目标，任何一个人工智能系统（包括机器翻译系统）的应对范围必须是“可界定的” (delimitible)，也就是必须是有限制的。为此，只有放弃机器翻译对象“无限制”这个目标才能够保证做到“心中有数”，实现翻译的可信赖性 (trustworthiness)，同时也才能将“可信赖性”置于系统设计的最高目标位置上。已经证明，为了实施可信赖机器翻译系统，必须放弃那些无法计算针对具体语言对象翻译正确性的信心指数 CFC 的翻译体系。

可信赖机器翻译系统的理论立场是：必须而且应该能够针对具体语言对象的翻译评估各自的 CFC，关键在于具体而有区别的评估，笼统的猜测对于我们是无意义的。

前几年，我们开发了某些子系统，如地名和单位名可信赖翻译系统等。在开发的过程中我们发现，只有在系统积累了大量具体而正确的翻译实例的时候，我们的翻译模型就自然而然地也就有了可信赖性，这个时候，翻译新的单位名的信心就自然而然地增强了。

所谓 CFC，是系统对各层次的具体语言对象翻译正确性的一个主观可认定、又客观可证明的度量。系统必须制订出对这些 CFC 的事先标记、事后证明与修正的方法。在此基础之上，TMT 的目标就可以实现了。同时，我们还可以通过不断修正的 CFC 实现自学习 (self-learning)，不断改进翻译能力。

所以，在我们提出可信赖机器翻译这个目标的同时，事实上也就公示了某些潜台词和潜在目标——就是要开发对多层次具体语言对象有评估翻译能力和学习能力的翻译系统。

本文证明，为了实现 TMT 或机器翻译的可信赖性，CFC based TMTSoLA 是不可避免的。

## 2 CFC及其数制

CFC (confidence factor of correctness)，即正确性信心指数，本来是任何 TMT 系统的一个核心概念，实现 CFC 的描述、计算和改进，是任何一个可信赖机器翻译系统必须解决的重大理论与工程问题。而由改进 CFC 所带来的学习能力 (Learning ability)，可以说是自然而然的结果。所以，我们将它叫做 CFC based TMTSoLA，它是 TMT 的实现形式。既然如此，它的基本要求就与 TMT 完全一样。

TMT 要能够对其所翻译的每一个具体的语言对象——源语言上下文里的每一个词语 (phrase)、句子 (sentence) 和上下文关系 (relation of context)——翻译的正确性做到心中有数。就是说，在给出翻译之前，必须为具体语言对象翻译的正确性打一个分数，这个分数就是 CFC。

TMT 系统一旦有了针对每一个具体语言对象翻译的可以查找计算的 CFC，也便能够随时将它自己可以正确翻译的部分和还不能正确翻译的部分区别开来。从而，TMT 将能够保证这些具体的语言对象翻译的可信赖性，而对于它没有把握的部分，要么加以翻译而标记出来要么将其留给别的系统翻译或者请教人来帮助。

正确性信心指数自然也是一个数。既然是一个数，从理论上讲，就可以是从二进制（binary）到任意分制的：如二分的（binary——有把握与无把握）、三分制的（有把握、两可、无把握）、五分制的（很有把握、有把握，两可、无把握，很无把握）、十分制的、百分制的和任意分制的等。显然，分制越高，系统对于“信心”将区分得越是精细。

所以，理论上，TMT 理论里的 CFC 确实有一个数制问题。而且还应该承认，有怎样的数制，就会有怎样的处理策略。所谓数制问题，其实也是一个核心的策略问题。

实际上，在各个数制里，最重要的还是二进制和百分制。让我们来研究两种分制引起的两种策略：一种是二进制 CFC 判定与处理策略，另一种是百分制 CFC 的判定与处理策略。

所谓二进制 CFC 判定与处理策略，其对正确性的信心指数只有“有把握”与“没有把握”两个值。自然，相应的策略是：它只翻译“有把握”的对象，凡是“没有把握”的，自然它就不加翻译了（留给别的系统去翻译或请人帮助）；对于相应的计算和翻译行为，也采取不是冻结就是激活的两极做法，所以一旦发现某一个层次的某个具体语言对象的翻译方法给出了错误的结果，将立即冻结该层次的该具体语言对象的翻译（模型）。

所谓百分制 CFC 策略，自然也要对各个层次的具体语言对象分别计算其 CFC 和具体地进行翻译，但是将该正确性的信心指数 CFC 用百分制加以度量，其正确性信心指数的数值在 0-100 之间。此时，它可以实行对该层次的全部语言实例加以翻译的做法，但是却能够在翻译之后标注出一个从 0-100 的信心指数来，0 为对其正确性毫无信心的，100 为对其正确性最有信心的。

若出现个别错误实例，百分制 CFC 策略首先是容忍，同时随时记录该层次具体的语言对象翻译正确与错误的次数，用它作为修正 CFC 的依据。这一点与 CF 的计算法似乎是一致的。这样的系统虽然表面上没有拒绝任何翻译，但是它提供的翻译结果却是心中有数，其对不同层次不同语言对象所提供的 CFC 显然是各不相同的。

百分制 CFC 策略可能有几个优点：

它不必因为某个层次的某个具体翻译模型出现一两个错误实例就手忙脚乱，从而减少被冻结的频率；

用实际翻译结果累计出来的正确率作为 CFC 的客观依据，比较有说服力；

解决了二进制 CFC 策略处理问题比较极端的嫌疑，提高了可信赖机器翻译理念的可接受性。

但是，这里绝无否定二进制 CFC 策略的意思。从社会心理学角度来看，二进制策略是多系统合作和交互学习过程中所必须的。为了能够合作，许多决定必须是两分的：“是是还是非”、“是对还是错”、“是有还是无”、“是好还是坏”、“是行还是不行”、“是同意还是不同意”、“是干还是不干”等等，几乎没有第三种选择的可能性。对于可信赖机器翻译而言，有正确把握与信心的就翻译，没有正确把握与信心的就拒绝翻译或者就发问，或者让给别的系统翻译或请人帮助。这样系统才可以进行有效操作。所以，两分制的 CFC 体系是一个必须的体系。

可是，正如人们在决定之前常常要考虑再三一样，为了正确地实现二进制 CFC 策略，我们也需要以百分制 CFC 策略作为决策的后台。若没有百分制 CFC 策略的支持，直接实行二进制 CFC 策略难免缺乏客观性，对于计算机来讲也将因过多的拒绝而使开发者与用户失去兴趣，增加其实现的难度。所以不妨使用诸如百分制或更灵活的实分制 CFC 策略，作为最后实现二进制 CFC 策略判定的基础。

### 3 CF、DF和CFC

前面已经提到过国内外机器翻译界长期以来已经习惯了的 CF（Correctness Factor）——翻译正

确率这个指标,我们来看它与CFC是不是有某种关系。

CF是得到正确翻译的句子数目除以全部句子数目计算出来的一个百分数:

$$CF = \frac{\text{得到本系统正确翻译的句子数目}}{\text{本系统翻译的句子数目}} \quad (1)$$

CF是在容许翻译模型存在错误的前提下计算出来的,具有一定的客观性和可计算性。因此,翻译正确率似乎可以被我们理解为对于一个翻译系统全部翻译结果正确性的一个客观度量。既然如此,那么,CF在TMT系统里有没有意义和有什么意义呢?比如,我们能不能利用翻译正确率作为TMT系统里的CFC呢?

众所周知,TMT允许拒绝翻译,而不允许在心中无数时进行有错误的翻译。所以CF与CFC没有任何关系,CF也根本不可以用做CFC。

对可信赖系统总体进行评价指标不再是翻译准确率CF,而是其可信赖度(DF)[2]。所谓可信赖度是一个TMT系统正确翻译的单位数目与总体数目的比值:

$$DF = \frac{\text{给出可信赖翻译的翻译单位数目}}{\text{翻译单位总数}} \quad (2)$$

实际上,“翻译正确率”这个指标是一个笼统的统计指标。既然是笼统的,所以它对不同的词语、不同的句子和不同的上下文关系没有兴趣。因此,CF是一个不可能运载具体个性的一个指标。

而用作具体词语、句子、上下文关系翻译正确性的信心指数(CFC)本身是以个性为基础的,与“翻译正确率”可以说是风马牛不相及的东西。

由于绝大多数的以“无限制、全自动”为目标的机器翻译系统的翻译模型是针对全部语言事实(无论是已知和未知的语言事实)的共同特性建立的,无论是正确的还是错误的翻译结果,都是在这个无所不包的模型支配下产生的,因此也就没有办法区别各个词语、各个句子、各个上下文关系翻译的CFC。因此,我们或计算机自然不可能总结出具有实际指导意义的改进模型的经验。可见,这样的评价体系和建立在该评价体系基础上的翻译模型,是导致不可信赖的翻译结果的根本原因。

若我们企图用一般机器翻译系统的翻译正确率来代替基于二分制CFC的TMT系统里的CFC,那么,CFC所特有的个性就不复存在了。

如果这样的话,对于一个明明已经得到了正确翻译的句子和一个明明不能正确地加以翻译的句子,都将给出完全相同的所谓的“CFC”来,这样所谓的“信心指数”根本谈不上是信心指数!这样,我们根本无法将不能翻译的句子识别出来、然后排除出去。所以,由“翻译正确率”来推算“CFC”这个指标压根儿就是错误的。

反过来讲,所有接受翻译正确率理念的翻译系统,从原理上来讲,哪一句会正确哪一句会错误它压根儿是心中无数的,所以没有能力判定哪一句是可以翻译的,哪一句是不可以翻译的。因而它挑选不出能够交给别的系统去翻译的东西。自然,也无法通过对于个别翻译的错误的发现来进行学习。

由此可见,对于采纳二分制CFC的TMT系统来讲,传统的翻译正确率是没有任何意义的。

#### 4 二分制CFC策略指导下的学习型可信赖翻译模型

二分制CFC策略指导下的TMTSoLA是TMT的一种形式,叫做“二分制CFC策略指导下的有学习能力的可信赖翻译系统”。

它不允许在给出的翻译里存在错误,它采取的是“一票否决式的管理方法”。就是说,在任何翻译模型里面,它不允许出现那怕是一个实例的错误。由任何独立的翻译模型指导下的翻译,一旦被发现出现一次错误,该模型将立即被冻结(freeze)或去活化(deactivite),只有把这个错误从该模型里排除出去之后,它才能够被解冻(thew-out)或重新被激活(reactivate)。

在这种情况下,排除该错误的最简单方法就是为每个模型附加一个反例口袋(BOE—bag of opposite examples)。一旦发现一例错误,只要将该错误实例加进反例口袋就算万事大吉。翻译模型工作之前,要首先检查反例口袋,若翻译对象与存放在反例口袋里的某个实例匹配的话,则拒绝翻译。这样,就可以消除该错误的再次发生,以保证该模型的可信赖性。

这样的基于反例口袋的学习方法是简单可行的。

当然,在这种学习方法驱使下,可信赖翻译模型的翻译效率和领地(territory)将随着反例口袋的扩大而逐步下降和缩小。但是,只要反例口袋是收敛的,该模型就是有效的。

百分制 CFC 策略指导下的学习型可信赖翻译模型

百分制 CFC 策略指导下的翻译模型同样需要为各个层次的具体的翻译对象建立各自的 CFC 算法和翻译算法,所不同的是,它不再施行“一票否决式的管理方法”。

百分制 CFC 策略是一种有限容错的策略,不必一发现有错就立即冻结相应的翻译模型。它也可以为每一个翻译模型建立一个反例口袋,进行自学习。但是,也可以不设反例口袋,而仅仅累计正确与错误翻译次数,不断计算特定模型基于累计的、类似于 CF 计算方法的 CFC。当某个模型的 CFC 低到某一个极限的时候,再加以冻结和处理。

百分制 CFC 策略的好处是避免过于频繁地冻结和修改模型,使得系统无所措手足,而且在模型改进之前能够为用户提供标记有不同的 CFC 的翻译结果。

## 5 如何为各个层次具体语言对象建立有效的CFC算法和可信赖的翻译模型

在可信赖机器翻译体系的设计里,能不能和如何在各个层次为具体的翻译对象建立和计算有效的 CFC 算法已经变为 TMT 的可行性的核心问题。

毫无疑问,传统的那种笼统的、不计算 CFC 的翻译模型充其量也只能作为一个候补模型,基本上不能被用作可信赖机器翻译系统的基本模型。可信赖机器翻译的模型必须保证 CFC 的具体化和个性:要让计算机在针对每一个或一类词语、每一个或一类句子、每一个或一类上下文关系的翻译模型里,留有一个确定的地方,在那里可以事先标记、事后修改翻译正确性的信心指数——CFC。

首先,可信赖机器翻译系统必须是一个建立在有针对性的 CFC 计算可能的模型基础上的翻译体系。针对每一个或每一类词语、每一个或每一类句子、每一个或每一类上下文关系,它必须有相对独立而具体的翻译算法和 CFC 的计算方法。这样看来,其翻译模型总数就不是决定性的,但是 CFC 所针对的具体对象和具体的翻译方法则肯定是大量的——按照我们现在的估计,应该在几十万到几千万之间。

其次,可信赖机器翻译系统在其针对每一个或每一类词语、每一个或每一类句子、每一个或每一类上下文关系的相对独立而具体的翻译方法的 CFC 计算算法下,需要一个空间和一个机制,来标记算法建立者事先估计的 CFC,同时还要为事后客观地记录翻译结果正确和错误的次数提供条件。这些客观记录是修改和证明事先关于 CFC 估计的客观基础。

第三,在实现前面这两点的基础之上,一种合作竞争机制将可以建立起来。按照竞争合作的原则,我们可以为每一个 TMT 子系统提供一个舞台。对于不同领地里的翻译对象各个子系统有各不相同的优先权,这个优先权是根据各自的领地大小(在二分制 CFC 策略下)或各自的 CFC(在百分制 CFC 策略下)大小决定的。这样,最后的竞争结果是各个子系统去翻译自己最有把握的部分,这样的系统的合作不仅仅是可信赖和有效的,而且,它对于人类的帮助也将是可以信赖的和有效的。

第四, 在实现前面这三点的基础之上, 一种经验型的学习机制将可以自动地建立起来。在此基础之上, 可信赖翻译系统将能够不断改正自己的翻译行为和翻译结果。比如, 前面提到反例口袋方法, 就类似于老鼠在寻找路径的过程中, 一旦发现某一块踏板烫脚就记住它, 以后就永远避开此踏板的学习算法。自然这一类学习方法比较笨, 因为这样一来, 系统的翻译效率越来越低, 翻译的领地也将越来越少, 但是, 这样却可以保证系统的可信赖性。

## 6 结论

对于任何可信赖机器翻译体系而言, 无论其翻译的基本方法是什么——是基于句法或语意规则的还是基于实例(语料库)的翻译, 都存在两种策略:

- (1) 是二分制CFC策略;
- (2) 是百分制CFC策略

对于任何可信赖机器翻译体系而言, 无论其翻译的基本方法是什么——是基于句法或语意规则的还是基于实例的(语料库)翻译, 无论是二分制CFC策略还是百分制CFC策略, 都必须做到下列三点:

- (1) 提供各个层次具体翻译对象特定翻译方法下的翻译正确性信心指数CFC的计算方法;
- (2) 提供存取这些信心指数函数所需要的参数的空间;
- (3) 根据人们对于其翻译结果的评价累计这些参数的方法。

## 参考文献

- 1 李应潭, 1999年4月, “人工智能的可信赖性及其算法研究”, 《信息与控制》
- 2 李应潭, 2000年5月, “TMT理论”, 研究报告, 中国科学院沈阳自动化研究所
- 3 Li Yingtan, 10/2001, “EQUIVALENT FORMULA OF S4 SYSTEM & ITS APPLICATION IN NLJ & TMT”, page 158, in *Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference* Copyright © 2001
- 4 黄昌宁, 董振东, 1999年10月, 《计算语言学文集》