

More than Words in Medical Question-and-Answer Sites: A Content-Context Congruence Perspective



合肥工业大学

Peng C H, Yin D, Zhang H. More than words in medical question-and-answer sites: a content-context congruence perspective[J]. Information systems research, 2020, 31(3): 913-928.

郝聪颖

2023.07

一 研究介绍

根据对1700名美国成年人进行的“美国医疗保健信息大搜索”调查，73%的人从互联网上获得健康相关信息。在另一项对1300多名美国人的全国代表性调查中，87%的青少年和年轻人（14-22岁）说他们在网上搜索过健康信息。



越来越多的患者依赖互联网作为他们在寻求专业诊断之前获得有关自身健康状况的知识的第一来源

最受欢迎的健康相关信息在线资源包括WebMD、Facebook、YouTube和Twitter。WebMD提供了问答(Q&A)平台，为了减少信息过载并突出更有价值的内容，允许读者对答案的有用性进行投票，并更突出地显示有用的答案。当未来读者决定阅读哪些内容时，很大程度上依赖于他人的判断。因此，更好地了解影响答案有用性的因素可以帮助未来读者更容易地找到有用的信息，并激励贡献者创造更多有用的信息。

Agarwal R, Gao G, DesRoches C, Jha AK (2010) Research commentary The digital transformation of healthcare: Current status and the road ahead. Inform. Systems Res. 21(4):796–809.

Cornejo C (2018) Social life of health information. Accessed July 16, 2019, <https://www.wegohealth.com/2018/02/05/social-life-of-health-information/>.

Stricker P 2014. Social media: An emerging communication modality. Accessed July 16, 2019, <http://www.naylornetwork.com/cmsatoday/articles/index-v2.asp?aid=296436&issueID=30171>.

— 研究介绍

在医学环境中，将对医学问题回答的有用性定义为该回答被认为有助于读者的健康相关判断和决定的程度。关于问答网站中答案的感知价值最常被研究的是答案长度或答案中包含的信息总量。此外，已经研究了围绕焦点内容的环境信号的各种背景线索的作用。（上下文线索是来自内容的环境信号）

我们的研究在两个重要方面与早期的工作不同：

- ✓ 首先，我们将研究定位在医疗问答网站的设置中，并强调两种独特类型的上下文线索的关键作用——问题的语言属性(具体性和情感强度)和与问题相关的疾病的严重性。
- ✓ 第二，尽管在理解内容特征和上下文线索对读者对答案的价值评估的影响方面取得了重大进展，但内容-上下文一致性的作用相对未被探索。

Huang L, Tan C-H, Ke W, Wei K-K (2013) Comprehension and assessment of product reviews: A review-product congruity proposition. J. Management Inform. Systems 30(3):311–343.

Yin D, Mitra S, Zhang H (2016) When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. Inform. Systems Res. 27(1):131–144.

二 理论与假设

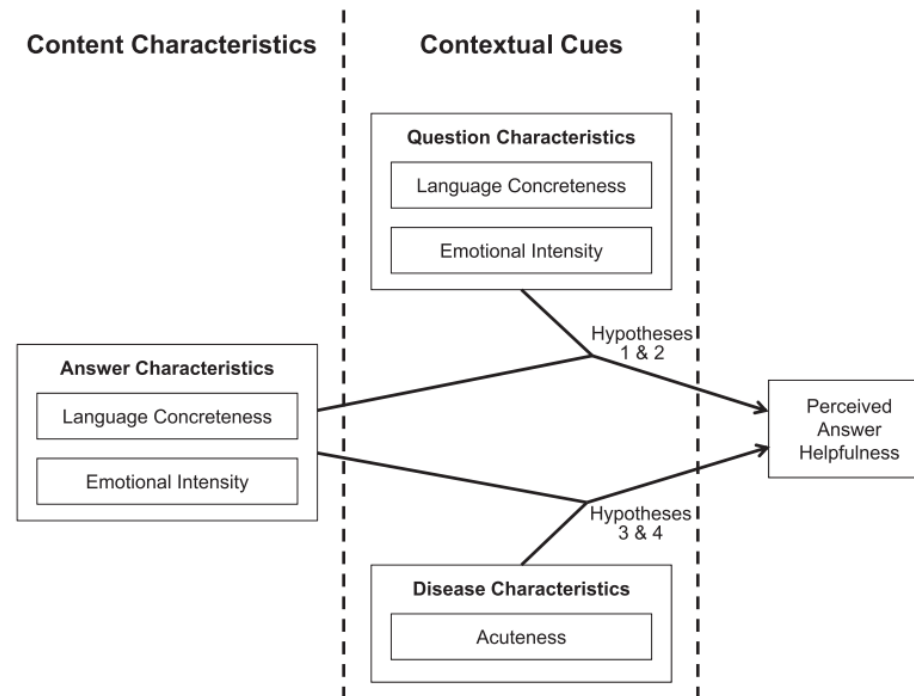
➤ 一致性

本文将文本一致性的概念应用于用户生成的内容，认为内容-背景一致性在读者的价值评价中起着不可或缺的作用。

- ✓ 答案内容的两个特点：语言的具体性和情感强度。
- ✓ 医学问答网站中特有的两类背景线索：问题内容的具体性和情感强度，以及疾病的严重性。

最后，提出了关于问题-答案一致性和疾病-答案一致性的假设。

Figure 1. Theoretical Framework



二 理论与假设

➤ 语言具体性与情感强度

答案贡献者可以以不同的方式编写答案表达相同的想法。除了消息的实质内容，其语言的属性(具体性和情感强度)也会影响受众。

在本文中，我们着重从两个方面探讨答题内容中所使用的语言：一个是认知的，另一个是情感的。

- 首先，语言的认知方面是它的具体性水平。将语言的具体性定义为词语提供关于物体或情境的描述性、具体性和生动性信息的程度。
- 除了内容的认知方面，答案贡献者也可能不同程度地表达感受。将情绪定义为内容贡献者表达的主观感受，并将情绪强度定义为内容中表达的情绪的百分比。

Berry DS, Pennebaker JW, Mueller JS, Hiller WS (1997) Linguistic bases of social perception. *Personality Soc. Psych. Bull.* 23(5): 526–537.

Bradac JJ, Bowers JW, Courtright JA (1979) Three language variables in communication research: Intensity, immediacy, and diversity. *Human Comm. Res.* 5(3):257–269.

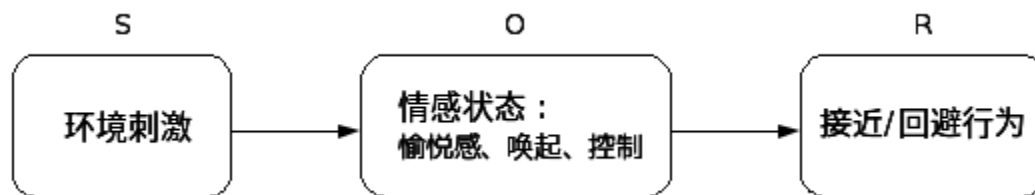
二 理论与假设

➤ 问题-答案一致性



答案内容可能在语言具体性和情感强度方面有所不同，而问题内容也可能在这些语言属性方面有所不同。问答一致性可能有助于感知价值的答案。

原因：刺激促进处理并增强流畅性。流畅性被广泛地定义为个人对他或她处理信息的轻松程度的主观体验。流畅性可以从问题的内容特征的刺激效应中产生。



二 理论与假设

➤ 问题-答案一致性

□ 语言的具体性

问题的写作方式可能会引导读者，使他们更流利地回答具有匹配内容特征的答案。

- 假设1.回答与问题之间的语言具体性一致性与回答的感知有用性呈正相关。

□ 情感强度

仅仅观察情感刺激，可以通过在观察者的记忆中预激活与情感相关的词和概念来引发观察者的情感。

此外，流畅性通常与积极评价相关联。流畅性通常用于识别更好的选择。应用到本文中，读者认为更流畅的答案应该得到更积极的评价。

- 假设2.答案和问题之间的情感强度一致性与答案的感知有用性呈正相关。

Yin D, Mitra S, Zhang H (2016) When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. *Inform. Systems Res.* 27(1):131–144.

Winkielman P, Huber DE, Kavanagh L, Schwarz N (2012) Fluency of consistency: When thoughts fit nicely and flow smoothly. Gawronski B, Strack F, eds. *Cognitive Consistency: A Fundamental Principle in Social Cognition* (Guilford Press, New York), 89–111.

二 理论与假设

➤ 疾病-答案一致性

不同疾病的一个特点是治疗时间不同，疾病通常被分类为持续有限时间段的急性疾病和持续时间较长的慢性疾病。

- ✓ 持续有限时间段的急性疾病：读者咨询往往会考虑如何及时治疗
- ✓ 持续时间较长的慢性疾病：读者往往会思考如何处理其长期后果，而这可能要到遥远的将来才能弄清楚

根据解释水平理论，时间距离对心理解释水平有直接影响。

时间距离和信息的具体水平之间的更大一致性可以使信息更有说服力，从而更具诊断性。这种效果可以用之前的流畅性参数来解释：当时间距离与信息的具体性一致时，这个人倾向于体验到更高的流畅性，这反过来又导致对信息的更积极的评价。

二 理论与假设

➤ 疾病-答案一致性

□ 语言的具体性

上述推理表明，疾病的严重性和信息的具体性之间的一致性可以增加感知的信息的有效性，而信息的具体性水平可以以不同的方式表现。

(1) 信息中使用的词语的具体性水平是信息具体性水平的最直接反映。医学环境中，当读者遇到对急性疾病问题的具体回答（或对慢性疾病问题的抽象回答）时，他们可能体验到更大的流畅性，并认为答案更有帮助。因此，提出以下疾病-答案的具体性一致性假设：

- 假设3.答案的具体性水平和疾病的严重性之间的一致性与读者对答案的感知有用性是正相关的

(2) 信息的具体程度可以体现在表达情感的强度上。调查数据表明情绪化的语言更具体和生动。当读者遇到与疾病严重性相匹配的情绪强度的答案时，会体验到更大的流畅性（对急性疾病问题的情绪回答，或对慢性疾病问题的非情绪回答），流畅性的积极体验导致读者对答案更积极的评价。因此，提出了以下疾病-答案情绪一致性假设：

- 假设4.答案的情绪强度与疾病严重性之间的一致性 与读者对答案的感知有用性 呈正相关。

三 方法与结果

➤ 数据

本文收集并分析了来自WebMD Answers的数据集。2018年1月，收集了所有26,318个问题及其答案的内容，分类为WebMD Answers创建的99个常见疾病主题。总共收集了35,098个答案，其中16,802个（超过46%）至少获得了一次投票。还收集了3022位提供答案的不同作者的数据。

三 方法与结果

➤ 测量

因变量是回答的有用性。对于来自WebMD Answers的答案，除了收集答案内容外，还收集了以下数据：“有用”的投票数和总投票数，用有用投票数与答案总投票数的比率来衡量有用性。

- 假设1.答案与问题之间的语言具体性一致性与回答的感知有用性呈正相关。
- 假设2.答案和问题之间的情感强度一致性与答案的感知有用性呈正相关。
- 假设3.答案的具体性水平和疾病的严重性之间的一致性与读者对答案的感知有用性是正相关的
- 假设4.答案的情绪强度与疾病严重性之间的一致性 与读者对答案的感知有用性 呈正相关。

自变量是四个认知和情感一致性变量：其中两个因素测量问题和答案之间的一致性 (H_1 和 H_2)，其他两个因素测量问题和答案中讨论的疾病主题的严重性之间的一致性 (H_3 和 H_4)。

在量化这些一致性变量之前，需要测量每个问题和答案的具体性和情绪强度，以及相关疾病的严重性。

三 方法与结果

➤ 测量

□ 内容的具体性

首先，为了衡量内容的具体性，采用通过互联网众包获得的近40,000个普遍已知的英语单词和表达的具体性评级字典。该字典包括来自医学领域的单词。例如，“消毒”具有2.67的具体性等级，而“主动脉”具有4.61的具体性等级。

用答案中单词的具体性评分总和 ÷ 答案中单词的总数计算答案的平均具体性。用类似的方法计算问题的平均具体性。

□ 情绪强度

其次，我们使用语言调查和字数统计 (LIWC) 软件来计算每个问题和答案的情绪强度。Pennebaker等人 (2007) 开发了一个心理测试验证的字典，由4,000多个单词和词干组成，分配给多个类别。该字典列出了表示积极或消极情绪的单词。我们通过计算情感词 (由LIWC字典识别) 的数量除以总单词数量来测量情感强度。

三 方法与结果

➤ 测量

□ 疾病的严重性

第三，通过家庭医生的专业评估来测量每个疾病主题的严重性。招募了三名美国家庭医生，每位医生阅读急性和慢性疾病的定义，然后沿着李克特五级量表评估每种疾病（1 =总是慢性，2 =经常慢性，3 =有时慢性有时急性，4 =经常急性，5 =总是急性）。为了评估该测量的可靠性，我们计算了每种疾病的三位医生的Cohen kappa系数，平均值为0.54，表明医生之间的一致性中等。因此，我们汇总了每种疾病的这些评级，并使用平均值作为疾病急性程度的衡量标准。

Cohen kappa系数是一个用于一致性检验的指标。计算是基于混淆矩阵的，取值为-1到1之间，通常大于0

A \ B	好	中	差
好	10	5	5
中	2	35	2
差	8	5	15

$$P_o = (10+35+15) / 87 = 0.689$$

$$a_1 = 10+2+8 = 20; a_2 = 5+35+5 = 45; a_3 = 5+2+15 = 22;$$

$$b_1 = 10+5+5 = 20; b_2 = 2+35+2 = 39; b_3 = 8+5+15 = 28;$$

$$P_e = (a_1b_1 + a_2b_2 + a_3b_3) / (87*87) = 0.455$$

$$K = (P_o - P_e) / (1 - P_e) = 0.4293578$$

三 方法与结果

➤ 测量

□ 四个一致性变量

最后，我们通过常用的差异评分法测量了四个一致性变量（因为在理论层面上，内容的具体性、情绪强度、疾病的严重性都独立于回答内容的有用性）

对于每个一致性变量，首先标准化两个分量。然后，计算这两个标准化分量之间的绝对差。将绝对差乘以-1来创建一致性变量的度量：值越高表示两个分量之间的一致性越大。

还设置一系列的控制变量，以解释来源、答案和问题特征的影响。

- 来源

对于Q&A设置，控制了答题者的专业知识，这是一个虚拟变量，表明答案的贡献者是否是专家。此外还考虑了答题者的可信度，用答题者收到的有帮助的投票总数除以答题者回答的问题总数。

三 方法与结果

➤ 测量

• 答案

首先，通过答案长度测量信息量。第二，答案阅读难度对其感知有用性有直接影响，故计算Coleman-Liau指数作为阅读困难的替代指标；第三，由于总投票数和内容有用性之间的潜在关系，将答案收到的总投票数作为对照；第四，较老的答案可能有更多的时间积累更多的选票，故控制了答案天数。第五，控制了答案的具体性和情绪强度；第六，前几个答案可能比后来的答案获得更多的投票，故控制了焦点答案的顺序。

• 问题

通过计算WebMD确定的问题关键字数量来说明问题的一般性。具有更多关键词被认为是更一般的问题，因为它可以从更多的角度回答，潜在地增加了创建有用答案的难度。此外，还将问题长度包括在模型中。

• 疾病

使用疾病水平固定效应来控制疾病主题异质性。

三 方法与结果

➤ 数据分析和结果

因变量，即有帮助的投票数与答案的总投票数之比，限制在0和1之间。鉴于有界的因变量会导致使用普通最小二乘法（OLS）的参数估计不一致，故使用双限制Tobit模型（分为观测模型和截断模型两部分）。

首先在模型1中输入控制变量；

然后在模型2中加入四个一致性变量。

Table 3. Tobit Regressions

	Model 1	Model 2
<i>Author expertise</i>	0.043*** (0.004)	0.041*** (0.004)
<i>Author credibility</i>	-0.001 (0.001)	-0.001 (0.001)
<i>Answer length</i>	0.026*** (0.001)	0.024*** (0.001)
<i>Answer reading difficulty</i>	0.004*** (0.001)	0.006*** (0.001)
<i>Answer total votes</i>	0.002 (0.001)	0.002 (0.001)
<i>Answer days</i>	0.049 (0.139)	0.038 (0.138)
<i>Answer concreteness</i>	0.004*** (0.001)	0.0001 (0.001)
<i>Answer emotional intensity</i>	0.003*** (0.001)	0.003** (0.001)
<i>Answer sequence</i>	0.006*** (0.001)	0.006*** (0.001)
<i>Other helpful answers</i>	0.002* (0.001)	0.002** (0.001)
<i>Question keywords</i>	-0.019*** (0.002)	-0.019*** (0.002)
<i>Question length</i>	-0.001 (0.002)	-0.001 (0.002)
<i>Question-answer concreteness congruence</i>		0.003*** (0.001)
<i>Question-answer emotion congruence</i>		0.003** (0.001)
<i>Disease-answer concreteness congruence</i>		0.006*** (0.001)
<i>Disease-answer emotion congruence</i>		0.004*** (0.001)
Constant	0.481 (0.412)	0.511 (0.411)
N	16,726	16,726
Log likelihood	5,519.81	5,558.64
p-value, LR test		0.000

三 方法与结果

➤ 数据分析和结果

假设1和2提出了问答具体性和情感强度的一致性 对 该答案的感知有用性的积极影响。

在表3的模型2中，问答具体性一致性与帮助性正相关 ($\beta = 0.003$, $p < 0.01$)，问答情感一致性与帮助性正相关 ($\beta = 0.003$, $p < 0.05$)。控制其他因素，一个答案的具体性或情感强度更接近其对应的问题，预计会更有帮助。因此，支持前两个假设。

Table 3. Tobit Regressions

	Model 1	Model 2
<i>Author expertise</i>	0.043*** (0.004)	0.041*** (0.004)
<i>Author credibility</i>	-0.001 (0.001)	-0.001 (0.001)
<i>Answer length</i>	0.026*** (0.001)	0.024*** (0.001)
<i>Answer reading difficulty</i>	0.004*** (0.001)	0.006*** (0.001)
<i>Answer total votes</i>	0.002 (0.001)	0.002 (0.001)
<i>Answer days</i>	0.049 (0.139)	0.038 (0.138)
<i>Answer concreteness</i>	0.004*** (0.001)	0.0001 (0.001)
<i>Answer emotional intensity</i>	0.003*** (0.001)	0.003** (0.001)
<i>Answer sequence</i>	0.006*** (0.001)	0.006*** (0.001)
<i>Other helpful answers</i>	0.002* (0.001)	0.002** (0.001)
<i>Question keywords</i>	-0.019*** (0.002)	-0.019*** (0.002)
<i>Question length</i>	-0.001 (0.002)	-0.001 (0.002)
<i>Question-answer concreteness congruence</i>		0.003*** (0.001)
<i>Question-answer emotion congruence</i>		0.003** (0.001)
<i>Disease-answer concreteness congruence</i>		0.006*** (0.001)
<i>Disease-answer emotion congruence</i>		0.004*** (0.001)
Constant	0.481 (0.412)	0.511 (0.411)
N	16,726	16,726
Log likelihood	5,519.81	5,558.64
p-value, LR test		0.000

三 方法与结果

➤ 数据分析和结果

假设3和4提出了疾病的严重性和答案的具体性或情感强度之间的一致性 对该答案的感知有用性的积极影响。在表3的模型2中，疾病-答案具体一致性与帮助性正相关 ($\beta = 0.006, p < 0.01$)，疾病-答案情感一致性与帮助性正相关 ($\beta = 0.004, p < 0.01$)。控制其他因素，在其具体性或情绪强度与其相应疾病主题的严重性之间具有较高水平的一致性的答案预计将更有帮助。因此，支持后两个假设。

Table 3. Tobit Regressions

	Model 1	Model 2
<i>Author expertise</i>	0.043*** (0.004)	0.041*** (0.004)
<i>Author credibility</i>	-0.001 (0.001)	-0.001 (0.001)
<i>Answer length</i>	0.026*** (0.001)	0.024*** (0.001)
<i>Answer reading difficulty</i>	0.004*** (0.001)	0.006*** (0.001)
<i>Answer total votes</i>	0.002 (0.001)	0.002 (0.001)
<i>Answer days</i>	0.049 (0.139)	0.038 (0.138)
<i>Answer concreteness</i>	0.004*** (0.001)	0.0001 (0.001)
<i>Answer emotional intensity</i>	0.003*** (0.001)	0.003** (0.001)
<i>Answer sequence</i>	0.006*** (0.001)	0.006*** (0.001)
<i>Other helpful answers</i>	0.002* (0.001)	0.002** (0.001)
<i>Question keywords</i>	-0.019*** (0.002)	-0.019*** (0.002)
<i>Question length</i>	-0.001 (0.002)	-0.001 (0.002)
<i>Question-answer concreteness congruence</i>		0.003*** (0.001)
<i>Question-answer emotion congruence</i>		0.003** (0.001)
<i>Disease-answer concreteness congruence</i>		0.006*** (0.001)
<i>Disease-answer emotion congruence</i>		0.004*** (0.001)
Constant	0.481 (0.412)	0.511 (0.411)
N	16,726	16,726
Log likelihood	5,519.81	5,558.64
p-value, LR test		0.000

四 结论与展望

• 结论

本文提出了一个内容与背景一致的观点，以解释医学问答网站上答案的感知价值，这补充了人们对背景线索在读者对某些内容的帮助性评估中的作用的兴趣。本文的研究提供了现实世界的证据，即如果答案内容与医学问答网站特有的语境线索相一致，那么它就会被评价为更有帮助。

• 展望

- ① 进一步探索内容与背景一致性对感知信息有用性的影响机制。当读者对某些内容进行理解时，他们的信念和心态在评价中起着关键作用。
- ② 探索提出的理论框架在其他类型的用户生成的内容中的普遍适用性（享乐主义/功利主义产品）
- ③ 内容-背景一致性的来源
- ④ 一个问题和一个答案在其他方面可能是一致的（情感价值），探索其他方面的一致性是否也促进答案的感知有用性
- ⑤ 在后两个假设中探索其他维度（如空间距离/社会距离）的影响，以及不同维度如何相互影响，从而影响用户的判断

信息鸿沟+不确定性

由于医疗的专业性和技术性等特点，医疗服务供需双方的信息不对称显著，医生在医疗服务市场拥有绝对信息优势，而患者医疗健康信息素养不足。Bardhan等人指出，理解和弥合健康信息系统和健康数据分析中的数字鸿沟（Digital Divide）是未来IS领域研究的一大挑战^[1]。

在患者端，帮助患者知晓自身情况和就医流程；在医生端，替代医生整理主诉、病史等关键信息。

《不仅仅是医学问答网站中的文字：内容-上下文一致性的观点》——与上下文内容和疾病严重性相匹配的答案对于患者有更大的感知价值

[1] Bardhan I, Chen H, Karahanna E. Connecting systems, data, and people: A multidisciplinary research roadmap for chronic disease management[J]. MIS Quarterly, 2020, 44(1): 185-200.

Q&A



合肥工业大学