

K sociologickým aspektům rozvoje vědy jako informačního systému

JAROSLAV KREJČÍ

Ústav pro filosofii a sociologii ČSAV, Praha

Povšimneme si některých otázek rozvoje vědy ze zorného úhlu teorie komunikace s přihlédnutím ke specifiku společenské vědy, zejména její experimentální části. Moderní teorie komunikace studuje systém komunikace jako stochastický nebo náhodný proces. Budeme tedy uvažovat o informačním modelu procesu rozvoje vědy s možnými aplikacemi v oblasti experimentálních společenských věd. Chce-li badatel v oblasti společenských věd konstruovat matematický model studovaného jevu, postupuje tak, že na základě logicky významných apriorních poznatků předchozího pokusu vybere jednu z několika konkurujících hypotéz. Hypotézy formuluje v matematickém jazyce a porovnává s experimentálními údaji. Výzkum končí výběrem té hypotézy, která je korektní s novými experimentálními údaji. [1]

Každá vědecká práce je tedy založena na jisté množině dříve formulovaných idejí. Nové vědecké práce jsou výsledkem využití bádání, která byla již dříve realizována, nebo jejich prohloubením. Věda je tedy adaptivní systém, jehož vývoj je řízen vnitřními informačními toky. Všechny vnější podmínky, tj. okolí systému, mohou mít na vlastní rozvoj vědy pouze sekundární vliv, jsou tedy buď akceleračními nebo retardačními činiteli. Je přirozené, že při studiu vědy jako informačního procesu vycházíme z růstu počtu vědeckých publikací, které pokládáme za nositele informací. [2]

1. Růst informačních toků ve vědě

Studujeme skupinu prvků — tzn. počet publikací, které mohou dát vzniknout se stejnou pravděpodobností dalším publikacím bez zániku. Předpokládáme, že v krátkém časovém intervalu h každá publikace s pravděpodobností $\lambda h + o(h)$ dává vzniknout nové publikaci. Kladná konstanta λ udává intenzitu vzniku nových elementů (publikací) a charakterizuje rychlost rozrůstání tohoto

souboru. Předpokládáme-li, že není žádná interakce mezi prvky a v okamžiku t je velikost souboru rovna n , pak pravděpodobnost zvětšení velikosti souboru v časovém intervalu $(t, t + h)$ je rovna $n\lambda h + o(h)$. Pravděpodobnost $P_n(t)$ toho, že velikost souboru je n , splňuje obecnou rovnici stochastického lineárního růstu pro $\lambda_n = n\lambda$. [3]

Tedy

$$\begin{aligned} P'_n(t) &= -\lambda_n P_0(t) + \lambda_{n-1} P_{n-1}(t) \quad (n \geq 1) \\ P'_0(t) &= -\lambda_0 P_0(t) \\ P_i(0) &= 1 \\ P_n(0) &= 0 \quad \text{pro } n \neq i \end{aligned} \quad (1)$$

Po dosazení $\lambda_n = n\lambda$ do (1) dostaneme:

$$\begin{aligned} P'_n(t) &= -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t) \quad (n \geq 1) \\ P_i(0) &= 1 \\ P_n(0) &= 0 \quad n \neq i \end{aligned} \quad (2)$$

Soustavu řešíme pomocí vytvořující funkce

$$G(s, t) = \sum_{n=0}^{\infty} P_n(t) s^n$$

ve tvaru

$$\begin{aligned} G(s, t) &= \left(\frac{s}{1-s} e^{-\lambda t} \right)^i = \\ &= s^i e^{-i\lambda t} \left[1 + \binom{i}{1} s(1-e^{-\lambda t}) + \right. \\ &\quad \left. + \binom{i+1}{2} s^2 (1-e^{-\lambda t})^2 + \dots \right] \end{aligned}$$

Tedy řešení pro $n \geq i$:

$$P_n(t) = \binom{n-1}{n-i} e^{-i\lambda t} (1-e^{-\lambda t})^{n-i} \quad (3)$$

Významnou pro naše úvahy je střední hodnota náhodné veličiny $r(t)$

$$(P_n(t) = P\{r(t) = n\})$$

1) $P'_n(t) = \lim_{h \rightarrow 0} \frac{P_n(t+h) - P_n(t)}{h}$

Střední hodnotu

$$M(t) = \sum_{n=0}^{\infty} n \cdot P_n(t)$$

získáme pomocí (2):

$$M'(t) = \lambda M(t)$$

$$M(t) = M(0) e^{\lambda t}$$

$$M(0) = i$$

Střední hodnota vyjadřuje proces exponenciálního růstu a je uváděna Nalimovem a Mulčenkem [4] jako popis mechanismu růstu počtu publikací. Využití Yule-Furryho procesu [3] pro model růstu informačních toků je tedy obecnějším přístupem, ovšem výhodným pro takové případy, kdy růst je ovlivňován nepředvídatelnými vlivy, které v mikrospecializacích jsou značné a významné. Vzhledem k předpokladu nulových interakcí mezi vznikajícími prvky je i tento popis hrubý a nepříliš vyhovující.

Dílčím zlepšením popisu informačního toku je zavedení sekundárních faktorů, vytvářejících překážky pro růst popsaný soustavou (2) a (3). Jde např. o jisté retardační faktory plynoucí z časově omezených disponibilních zdrojů. Shrňme-li všechny tyto faktory do jednoho brzdícího faktoru $L > n$ můžeme

pro brzděný růst psát $\lambda_n = n\lambda \left(t - \frac{n}{L}\right)$

a po dosazení do (1) máme

$$P'_n(t) = -n\lambda \left(t - \frac{n}{L}\right) P_n(t) + (n-1)\lambda \cdot$$

$$\cdot \left(t - \frac{n-1}{L}\right) P_{n-1}(t)$$

$$P_r(0) = 1$$

$$P_n(0) = 0 \quad \text{pro} \quad r \neq n$$

Po úpravách analogických předchozím dostáváme parciální diferenciální rovnici

$$\frac{\partial G}{\partial t} = \frac{\lambda s(s-1)}{L} \left[(L-1) \frac{\partial G}{\partial s} - s \frac{\partial^2 G}{\partial s^2} \right]$$

s podmínkou $G(s, 0) = s^r$ pro vytvořující funkci

$$G(s, t) = \sum_{n=0}^{\infty} P_n(t) s^n$$

Hledat řešení této diferenciální rovnice není nadějně. V publikaci [7] je uveden možný přístup k nalezení alespoň dílčích výsledků.

Autoři v monografii [4] shrnují řadu dříve publikovaných případů, dokládajících empirickou platnost výše uvedených závěrů, a to jak v makrosféře, tj. v jednotlivých základních vědních odvětvích, tak také v mikrosféře, tj. v dílčích specializacích jednotlivých odvětví. Je zřejmé, že model nám dává možnost charakterizovat růst informačního toku řadou parametrů, k nimž je dobré uvést pravděpodobnou rychlost růstu toku, pravděpodobnou velikost souboru v časovém okamžiku t a tím tedy i umožňuje pokus o nákladové ohodnocení. Například spočítaná pravděpodobná rychlost růstu informačního toku v dílčích specializacích vědeckého bádání může být přijata jako operacionalisticky definovaná míra aktuálnosti příslušného vědního směru, a tedy koneckonců i atraktivnosti jeho rozvoje.

2. Studium prostředků sloužících jako informační kanály v rozvoji vědy

Pokus o aplikaci teoreticko-informačního přístupu při výzkumu prostředků sloužících jako informační kanály v rozvoji vědy vede záhy ke značným těžkostem. Souvisí to s tím, že shannonovská míra neurčitosti závisí pouze na pravděpodobnosti různých výsledků experimentu,² nezávisí však na tom, jaké jsou vlastně tyto experimenty samy, jsou-li v jistém smyslu sobě „blízké“, nebo „vzdálené“. Například je-li příjemci dána informace, která je mu známa, pak přírůstek jeho „subjektivní“ informace je nulový.

Bude nutné změnit klasickou definici míry informace a ohodnotit význam přijaté informace pro příjemce. K tomu využijeme výsledků Weissovy práce [8], které pro přehlednost dále uvedeme v míře nezbytné pro naše potřeby.

Mějme neprázdnou množinu Q publikací majících z hlediska příjemce následující vlastnosti:

1. Q obsahuje nezávisle na příjemci jistou „neurčenou“ zprávu, značenou dále jako Θ .

2. Q je uspořádaná. Pro libovolné $e_1, e_2 \in Q$, pro které neplatí současně $e_1 = \Theta, e_2 = \Theta$, platí $e_1 \geq e_2$ právě tehdy, je-li číselná veličina přiřazená zprávě e_1 větší nebo rovna číselné veličině přiřazené zprávě e_2 . Vztah $\geq$ je tranzitivní na Q . Pro libovolné dvě zprávy e_1 a e_2 platí buď $e_1 \geq e_2$ nebo $e_1 \leq e_2$. Jestliže $e_1 \leq e_2, e_1 \geq e_2$, pak zprávy jsou identické a píšeme $e_1 = e_2$.

² Termínu experiment je použito v obvyklém pravděpodobnostním smyslu.

3. Necht $e_n, e_0 \in Q$; $e_0 \geq e_n, e_0 \neq e_n$. Pro libovolné $e_a \in Q, e_0 \geq e_a \geq e_n$ a libovolného příjemce B existuje číslo $n_a = n(e_a, e_0, e_n; B)$, $0 \leq n_a \leq 1$ takové, že pro B zpráva e_a je rovnocenná výsledku hry, ve které e_0 je realizována s pravděpodobností n_a a e_n s pravděpodobností $1 - n_a$, zapíšeme to ve tvaru $e_a \tilde{B} [n_a e_0, (1 - n_a) e_n]$.

4. Platí následující substituční pravidlo:

Je-li $E \geq e_0 \geq e_a \geq e_n \geq e$ a dále

$$e_a \tilde{B} [n_a e_0, (1 - n_a) e_n]$$

$$e_0 \tilde{B} [v_0 E, (1 - v_0) e]$$

$$e_n \tilde{B} [v_n E, (1 - v_n) e]$$

pak $e_a \tilde{B} [n_a e_0, (1 - n_a) e_n] \tilde{B} [(n_a(v_0 - v_n) + v_n) E, (n_a(v_n - v_0) + (1 - v_n)) e] = [v_a E, (1 - v_a) e]$

5. Zprávy $e \in Q_1, f \in Q_2$ jsou „ekvivalentní“ vzhledem k příjemci B ; $e \tilde{B} f$, jestliže B nemůže preferovat jednu z těchto zpráv před druhou (neurčené zprávy Θ jsou vždy ekvivalentní).

6. Necht $e_0, e_a, e_n \in Q_1; f_0, f_b, f_n \in Q_2; f_0 \geq f_b \geq f_n$. Jestliže $e_0 \tilde{B} f_0, e_n \tilde{B} f_n$ a $e_a \tilde{B} [n_a e_0, (1 - n_a) e_n], f_b \tilde{B} [v_b f_0, (1 - v_b) f_n]$ pak $e_a \tilde{B} f_b$ tehdy a jenom tehdy, když $n_a = v_b$.

7. Je-li $e_a \tilde{B} [n_a e_0, (1 - n_a) e_n]$ a $e_a \tilde{B} f_b$, pak $[n_a e_0, (1 - n_a) e_n] \tilde{B} f_b$ a obráceně.

Oblast zájmů příjemce B , která je totožná se systémem možných typů relevantních publikací, označme I . Tedy I je systém množin $Q_i, I = \{Q_1, \dots, Q_m\}$. Oblast zájmů I bude mít následující vlastnosti.

8. Systém $I = \{Q_1, \dots, Q_m\}$ je maximální pro B v tom smyslu, že pro všechna $e \in Q \notin I$ a $\Theta \in Q_i, i = 1, 2, \dots, m$ vždy $e \tilde{B} \Theta$, přičemž $I' = \{Q_{i1}, Q_{i2}, \dots, Q_{im-1}\}$, kde $Q_{is} \neq Q_{ie}; s \neq e; Q_{is} \in I$ není oblastí zájmu B .

9. Pro libovolné $Q_i \in I$ existuje zpráva závislá na příjemci $B, e_{0i} \geq \Theta$ taková, že pro libovolné $E \in Q_i$, pro něž $E \geq e_{0i}$, platí $E \tilde{B} [1 \cdot e_{0i}, 0 \cdot \Theta]$ (e_{0i} je hodnota „nasycení“, příjemce).

Necht je dán experiment A (realizace vědecké práce) s výsledky $A_1, A_2, \dots, A_n$, jejichž pravděpodobnosti jsou rovny $p_1, p_2,$

$\dots, p_n, \sum_{i=1}^n p_i = 1$. Experiment A skuteční příjemce B , přičemž B má jistou hypotézu o výsledcích experimentu. Předpokládá, že výsledky A_i mají (subjektivní) pravděpodobnosti $q_1, q_2, \dots, q_n; q_i \geq 0; \sum_{i=1}^n q_i = 1, q_i \neq p_i$.

Na systému $I = \{Q_1, \dots, Q_m\}$ existuje jediný prvek Q_{i0} takový, že hodnota e_{0i0} je maximální. To znamená, že pro všechna $Q_j \in I$ existuje $n_j, 0 \leq n_j \leq 1$ takové, že

$$e_{0j} \tilde{B} [n_j e_{0i0}, (1 - n_j) \Theta] \quad (4)$$

přičemž $n_{i0} = 1$.

Pro každé A_i a každé $Q_j \in I$ utvoříme $A_i | Q_j = e_{ij} \in Q_j$ a definujeme číslo $n_{ij}, 0 \leq n_{ij} \leq 1$, pro které

$$e_{ij} \tilde{B} [n_{ij} e_{0j}, (1 - n_{ij}) \Theta] \quad (5)$$

Z vlastností 4, 6 a 7 plyne, že

$$e_{ij} \tilde{B} [a_{ij} e_{0i0}, (1 - a_{ij}) \Theta] \quad (6)$$

kde $a_{ij} = n_{ij} \cdot n_j$ a proto $0 \leq a_{ij} \leq 1$.

Máme nyní možnost zavést pojem subjektivní neurčitosti experimentu. Je třeba si uvědomit, že výsledky A_i jsou statisticky nezávislé, ovšem pro příjemce informace B mezi nimi existuje „shoda“, nebo jak jsme v úvodu odstavce říkali „blízkost“. S cílem formalizace pojmu shody nebo blízkosti zavedeme metriku:

$$d^{(j)}(A_i, A_k) = 1 - |a_{ij} - a_{kj}| = d_{ik}^{(j)} \quad (7)$$

Subjektivní neurčitost vzhledem k $Q_j \in I$ budeme nyní definovat jako

$$H_s^{(j)} = - \sum_{i=1}^n a_{ij} p_i \lg \sum_{k=1}^n d_{ik}^{(j)} q_k \quad (8)$$

a subjektivní neurčitost experimentu A vzhledem k příjemci B jako

$$H_s = \sum_{j=1}^m H_s^{(j)}$$

Analogicky je možné zobecnit pojem informace definicí subjektivní informace. Necht $H_s(A/M)$ je podmíněná subjektivní neurčitost experimentu A vzhledem k příjemci B za podmínky, že B získá nějakou zprávu M . Zpráva M může vést ke změně objektivních

a_{ij} nebo subjektivních q_i pravděpodobností výsledků experimentu A nebo ke změně a_{ij} . Ve stavu „uspokojení“ příjemce B se některá a_{ij} zmenšují, ve stavu „neuspokojení“ rostou. Ukazuje se rozumným přijmout

$$I_s(M|A) = H_s(A) - H_s(A|M) \quad (9)$$

jako míru subjektivní informace získané příjemcem B , dostane-li zprávu M týkající se experimentu A .

V odstavci 1 jsme odvodili některé závěry týkající se kvantifikaci informačního toku a naznačili možnosti plynoucí z toho pro objektivní sledování atraktivnosti jednotlivých odvětví a tím i mimo jiné k získání objektivních údajů pro preferenční škálu. Odstavec 2 dává naopak možnost sledovat efektivitu vědecké práce individua v nejširším smyslu.

3. Možnost definice řízení v informačním systému vědy

Tato závěrečná poznámka má pouze naznačit možnost zavedení procesu řízení na modelu informačního systému vědy. Formulujeme úlohu obecně. Sledujeme prvky (publikace), které se vyvíjejí navzájem jen nezávisle a dávají vzniknout další generaci publikací. Každá publikace je v další generaci nahrazena n_1 publikacemi typu Q_1 , n_2 publikacemi typu Q_2 , ... n_r publikacemi typu Q_r . Pravděpodobnost přechodu pro typ Q_i je $p(Q_i; n_1, n_2, \dots, n_r)$. Způsob ovlivnění (řízení) bude vyjádřen parametry I_s odvozenými v (9). Předpokládejme, že jde o časově homogenní řízení, tzn. parametr I_s nezávisí na generaci. Proces, který je popsán soustavou rozložení pravděpodobností,

$$\{p(Q_i, n_1, n_2, \dots, n_r; I_s(Q_i)), n_j \in \gamma; j = 1, \dots$$

$$\dots r\}$$

$$Q_i \in I,$$

$$I_s(Q_i) \in I_s$$

(γ — číselná množina)

je řízený Galtonův-Watsonův větvičí proces. I je systém možných typů relevantních publikací definovaný v odstavci 2.

Mějme dále výnos $C(Q_i, n_1, n_2, \dots, n_r; I_s)$, $Q_i \in I$, $n_j \in \gamma$, $j = 1, \dots, r$, z publikace typu Q_i při ovlivnění $I_s(Q_i)$ (dále pouze $I_s(i)$). Očekávaný výnos z jedné publikace typu Q_i a z $N - 1$ generací jejich potomků při ří-

zení I_s označme $\mu^N(Q_i, I_s)$. Očekávaný výnos z publikace typu Q_i bude

$$r(Q_i, I_s) = \sum_{n_1, \dots, n_r} p(Q_i; n_1, \dots, n_r; I_s) \cdot C(Q_i; n_1, \dots, n_r; I_s)$$

a očekávaný počet potomků typu Q_i této publikace je

$$m(Q_i, \overline{Q_j}, I_s) = \sum_{n_1, \dots, n_r} p(Q_i; n_1, \dots, n_r; I_s) n_j$$

Platí rekurentní vztah

$$\mu^N(Q_i; I_s) = r(Q_i; I_s(i)) + \sum_j m(Q_i, Q_j, I_s(i)) \cdot \mu^{N-1}(Q_j; I_s)$$

Maximální výnos $\max_{I_s} \mu^N(m_i, I_s)$, vyhovuje

Bellmanově rovnici.

Lze odvodit metodu k nalezení takového I_s , pro něž je růst populace nejrychlejší. [11]

Využijeme-li tedy Galtonova-Watsonova větvičího procesu pro modelování populace publikací, bude řídicím parametrem subjektivní informace I_s , která bezprostředně ovlivňuje výběr stavových přechodů, charakterizovaných příslušným rozložením. Výnosová f-ce je pak ohodnocením očekávaného počtu publikací v n -té generaci při řízení I_s . Očekávaný výnos (počet publikací) tedy získáme jako součin přechodové f-ce a výnosu jedné publikace v n -té generaci při řízení I_s . Maximálnímu výnosu bude tedy odpovídat optimální trajektorie stavových přechodů a optimální řídicí parametr I_s .

Celá praktická aplikace si vyžádá konkrétní formulace na statisticky získaných podkladech. Tematicky blízký a podstatně obsáhlejší přístup k řešení dané problematiky je obsažen v pracích M. Kochena [5] a [6].

Literatura

- [1] František Charvát, K technice tzv. inverzního dotazníku při obsahové analýze verbálních dokumentů, referát na VII. svět. sociolog. kongresu ve Varně.
- [2] Derek J. Price, *Little Science, Big Science*, Columbia University Press N.Y. 1963.
- [3] W. Feller, *An Introduction to Probability Theory and its Applications* I. 1957 N.Y.
- [4] V. V. Nalimov, Z. M. Mučenko, *Naukometria*, Nauka, Moskva 1969.
- [5] Kochen M., *The Growth of Knowledge*, New York, Wiley 1967.
- [6] Kochen M., *Stability in the Growth of Knowledge*, Michigan, Univ. Press 1969.
- [7] Barlett M. S., *An Introduction to Stochastic Processes*, N.Y. 1955.

- [8] Weiss P., *Subjektive Unsicherheit und subjektive Information*, Kybernetik, 1968, 5, No.2, 77—82.
- [9] Harris T. E., *The Theory of Branching Processes*, Berlin, Göttingen, Heidelberg 1963.
- [10] Boltjanskij V. G., *Matematičeskije metody optimalnogo upravljenja*, Moskva 1969.
- [11] Mandl P., *Řízené Markovovy řetězce*, Příloha časopisu Kybernetika, ročník 5/1969.

Резюме

Крейчи И.: К социологическим аспектам развития науки как системы информации

В статье изучается информационная модель процесса развития науки в связи с возможными приложениями в области общественных наук. Наука понимается как адаптивная система, развитие которой управляется внутренними потоками информации. Влияние окружения системы принимается во внимание только во второй аппроксимации. Анализ информационного процесса исходит из роста числа научных публикаций, которые считаются носителями информации. Для описания роста приводятся разные типы моделей и сравнивается их адекватность. Есть ссылка на возможность использования характеристик потока информации в определении меры актуальности научной специализации.

При изучении средств, служащих информационными каналами, использовано понятие субъективной информации, введенное Вейсом. Показана возможность применения субъективной информации при наблюдении эффективности научной работы. В заключение приводится описание определения процесса управления в этой системе. В качестве управляющего параметра использованы субъективные информации. Функция получения тогда является оценкой ожидаемого числа публикаций в n -ой генерации при данном управляющем параметре. Ожидаемое получение приобретает как произведение переходной функции и получения одной публикации в n -ой генерации при данном управляющем параметре. Максимальному получению соответствует оптимальный управляющий параметр.

В статье приводятся некоторые заключения, касающиеся квантификации потоков информации, и намечены возможности, вытекающие из этого для объективного наблюдения привлекательности отдельных открытий и, помимо прочего, приобретения объективных данных для шкалы предпочтений. Далее следует за возможностью измерения эффективности научной работы индивидуума. Наконец приводится по-

пытка ввести процесс управления в исследуемую информационную систему.

Summary

Krejčí J.: On Sociological Aspects of Development of Science as an Informative System

The article deals with the informative model of the process of science development in connection with possible applications in the sphere of social sciences. Science is understood as an adaptive system, the development of which is controlled by inner information channels. The influence of the environment of the system is taken into consideration in the second approximation only. The analyses of the information process is based on the growth of number of science publications, which are being considered as source of information. For description of growth there are stated different types of models and their adequateness is compared. It is shown that there is a chance of utilisation of characteristics of informative course in definition of actuality measure of growth of science specialization.

When studying the means serving as informatory channels in science development, the concept of subjective information derived by Weiss is made use of. The author describes a possibility of use of subjective information, when studying effectivity of scientific work and defines the control process of this system. Subjective information is made use of as a control parameter. Gain function then means evaluation of expected number of publications in the n -generation at a given control parameter. The expected gain is the product of the transitional function and gain of one publication in the n -generation at given managing parameter. The maximum gain corresponds with the optimum managing parameter.

Some conclusions concerning the quantification of information channels are derived and possibilities arising from this are indicated; these facilitate the objective investigation into the attractiveness of separate discoveries and thus, among other things, also the acquisition of objective data for the preference scale. Further, the possibility of measuring the effectivity of the individual's scientific work is examined, and finally, an attempt at introducing the process of management into the studied information system is presented.