

LINGVISTIK ANNOTATSIYA ISHONCHLILIGINI BAHOLASH: KO‘P QATLAMLI TAHLIL METODLARI

Daminova Faridaxon Bahriddin qizi

daminovafaridaxon@gmail.com

Guliston shahar 6-maktab o'qituvchisi

Annotatsiya. Mazkur maqola ko‘p qatlamli lingvistik annotatsiyaning ishonchliligini baholash metodlariga bag‘ishlangan. Annotatorlararo kelishuvni bitta umumiy foizga qisqartirish qaysi komponent – hodisaning mavjudligi, span chegarasi yoki kategoriya tanlovi – ishonchsiz ekanini yashirib qo‘yadi. Shu sababli maqolada uch mustaqil o‘lchov ko‘rib chiqiladi: Cohen's κ (nominal kategoriya kelishuvi), Krippendorff's α (bir nechta annotator va yetishmayotgan belgilar uchun umumlashgan yondashuv) hamda erkin span vazifalari uchun γ va exact-match F1 ko‘rsatkichlari. Har bir o‘lchovning qo‘llash sharti, matematik asosi va talqin chegaralari tahlil qilinadi, shuningdek kamyob sinflar va «prevalence paradox» kabi metodik xavflar, ikki annotatorli pilot sxema va ma'lumotlarni validatsiyalash protokoli taqdim etiladi.

Kalit so‘zlar: *annotatorlararo kelishuv, Cohen's kappa, Krippendorff's alpha, gamma (γ) o‘lchovi, exact-match F1, annotatsiya ishonchliligi, pilot annotatsiya, oltin standart.*

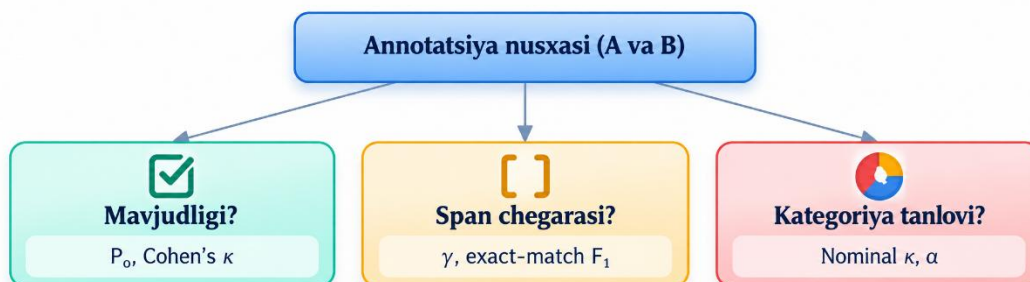
Kirish: nega bitta koeffitsiyent yetarli emas

Lingvistik annotatsiyaning ishonchliligi ma'lumotni boshqa annotator qay darajada o‘xshash belgilashi bilan bog‘liq, biroq har xil vazifa bitta koeffitsiyentda to‘liq aks etmaydi. Annotatorlararo kelishuv metodlari annotatsiya birligi, kategoriya tabiati hamda ehtimoliy tasodifiy moslikka sezgirligini hisobga olishi kerak. Ko‘p qatlamli annotatsiya sxemasida, xususan, uchta tubdan farqli qaror alohida-alohida o‘lchanadi:

1. Nomuvofiqlikning mavjudligini ajratish – birlik umuman nomuvofiqlikmi, yo‘qmi?
2. Ajratilgan matn oralig‘ining (span) chegaralarini moslashtirish – ikki annotator bir xil joyni belgilaydimi?
3. Bir xil birlik uchun IML/USL/TER va ichki tegni tanlash – qaysi lisoniy hodisa deb topildi?

Uch mustaqil savol, uch xil o'lchov talab qiladi

Bitta umumiy foiz bu farqni yashiradi



1-rasm. Ko'p qatlamli annotatsiya sxemasi

Agar bu uch bosqich bitta umumiy foizga yig'lsa, aynan qaysi komponent ishonchsiz ekani ko'rinmay qoladi: yuqori umumiy kelishuv aslida past span mosligini yoki kategoriya tanlovidagi nomuvofiqlikni yashirishi mumkin. Shu sababli quyida har bir komponent uchun mos o'lchov alohida ko'rib chiqiladi.

Cohen's kappa: nominal kategoriya kelishuvi

Cohen's κ ikki annotatorning bir xil, oldindan ajratilgan nominal birlikka qo'ygan kategoriya hukmini baholashda qo'llaniladi. Bu o'lchovda tasodifiy moslik P_e annotatorlarning kategoriya ulushlari ko'paytmalari yig'indisidan aniqlanadi:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

$$P_e = \sum_c p_a(c) \cdot p_b(c)$$

Bunda P_o – kuzatilgan moslik, P_e – annotator sinf ulushlaridan kelib chiqadigan tasodifiy moslik. Biroq ikki annotator matn oraliqlarini mustaqil ravishda ajratgan bo'lsa, ularning span ro'yxatlariga to'g'ridan-to'g'ri κ hisoblash matematik va metodik jihatdan asoslanmaydi – buning uchun avval umumiy qaror birliklari aniqlanadi yoki spanlarni moslashtirishga ixtisoslashgan o'lchovlar tanlanadi (keyingi bo'limga qarang).

Krippendorff's alpha: ko'p annotator va yetishmayotgan belgilar

Krippendorff's α bir nechta annotator ishtirok etishi, ayrim birliklarda belgi yetishmasligi yoki kategoriya masofasini turli tarzda belgilash kerak bo'lganda qulay umumiy yondashuvlardan biridir. Uning asosiy shakli kuzatilgan kelishmovchilik D_o ning kutilgan kelishmovchilik D_e ga nisbati orqali ifodalanadi:

$$\alpha = 1 - \frac{D_o}{D_e}$$

Biroq “ α hisoblandi” degan gapning o‘zi mazmunli natija emas: qanday birliklar, qaysi masofa funksiyasi (masalan, nominal) va qanday yetishmaydigan belgilar qoidasi qo‘llangani ochiq yozilishi lozim. Taklif etilayotgan protokolda IML, USL va TER qatlamlari uchun α ni alohida hisoblash, umumiy teglar soni katta bo‘lganda esa kam uchraydigan sinflarni ham alohida tavsiflash tavsiya etiladi.

Span kelishuvi: gamma yondashuvi va exact-match F1

Span chegarasi matnni «unitizing» qilish vazifasidir: annotatorlar nafaqat sinfni, balki hodisa qayerda boshlanib tugashini ham mustaqil aniqlaydi. Bunda aynan bir xil chegara, qisman kesishuvchi chegara yoki butunlay mos kelmagan hodisalar mavjud bo‘ladi. Mathet, Widlöcher va Métivier tomonidan taklif etilgan γ yondashuvi annotatsiya birliklarini moslashtirish bilan kelishuvni baholashni birgalikda ko‘rib chiqadi; erkin span vazifalari uchun birlik chegarasini hisobga oladigan bu kabi usulni qo‘llash imkoniyati mazkur protokolda alohida tekshiriladi.

Oddiy va tushunarli qo‘shimcha diagnostika sifatida spanlararo exact-match precision, recall hamda F_1 ham ishlatilishi mumkin:

$$F_1(\text{span}) = \frac{2 \cdot P(\text{span}) \cdot R(\text{span})}{P(\text{span}) + R(\text{span})}$$

Bunda A annotator spanlari B annotatorning mos spanlariga qarshi solishtiriladi; ikki tomon teng huquqli annotator bo‘lsa, natija A/B almashlab qayta hisoblash yoki simmetrik moslik orqali tekshirilishi mumkin. Exact-match F_1 tasodifiy kelishuvni tuzatmaydi va κ yoki α ning o‘rnini to‘liq bosa olmaydi – u aynan qayerda chegara farqi yuzaga kelganini ko‘rsatadigan yordamchi ko‘rsatkich sifatida talqin qilinadi. Qisman mos kelgan spanlar esa alohida statistikada berilib, aniq moslik bilan aralashtirilmaydi.

Ko‘rsatkichlar tanlovi va qo‘llash cheklovlari

Yuqorida ko‘rib chiqilgan uch o‘lchov har xil savollarga javob beradi va biri ikkinchisining o‘rnini bosmaydi. Quyidagi jadval qaysi baholash savoliga qaysi ko‘rsatkich mos kelishini va uning qo‘llash shartini umumlashtiradi.

1-jadval. Annotatsiya sifatini baholash ko‘rsatkichlari, ularni qo‘llash shartlari va metodologik cheklovlar

Baholash savoli	Taklif etilgan ko'rsatkich	Qo'llash sharti / ehtiyot
Birlikda xato bormi?	P_o va Cohen's κ	Bir xil oldindan belgilangan birliklar
Asosiy IML/USL/TER tegimi?	Nominal κ ; α	Kelishilgan o'lchov birligi; sinf taqsimoti
Ko'p annotator / ayrim yetishmagan belgilar	Krippendorff's α	Masofa, missing-data qoidasi ochiq beriladi
Erkin tanlangan spanlarning chegarasi	γ yoki segment mosligi diagnostikasi	Unitizing va alignment protokoli zarur
Span aynan bir xil topildimi?	Exact-match span F_1	IAA o'rnini to'liq bosmaydi
Kamyob ichki teglar	Sinf chastotasi; chalkashish jadvali	Faqat umumiy o'rtacha bilan xulosa qilmaslik
Janr-soha umumlashuvi	Qatlamlangan qiymatlar, ishonch oralig'i	Hujjat oilalari bo'yicha bog'liqlik nazorati

Izoh: κ , α , γ yoki F_1 bo'yicha ushbu bo'limda real koeffitsiyent hisoblanmagan; jadval analitik o'lchash protokolining loyihasidir.

Koeffitsiyentlarni talqin qilish: kamyob sinflar va “prevalence paradox”

κ yoki α qiymatini bitta qat'iy universal kesish chegarasi bilan “muvaffaqiyatli” yoki “muvaffaqiyatsiz” deb e'lon qilish metodologik jihatdan ehtiyotkorlikni talab etadi. Vazifaning subyektivlik darajasi, teglar soni, kam uchraydigan sinflar, annotatorlar tayyorgarligi va matn birligi koeffitsiyentga sezilarli ta'sir qiladi. Shu sababli loyiha uchun maqbul sifat chegaralari pilot natijalar, lingvistik xatar va keyingi modellarning maqsadli qo'llanishi bilan asoslanishi lozim; nazariy matnda allaqachon olingan qiymatlarni natija deb taqdim etish mumkin emas.

κ ga “prevalence paradox” deyiladigan holat ta'sir qilishi mumkin: kuzatilgan moslik yuqori bo'lsa ham, ayrim sinflarning juda notekis taqsimoti kappas (κ) qiymatini kamaytirishi yoki talqinni murakkablashtirishi mumkin. Shuning uchun koeffitsiyent yonida oddiy moslik, sinf ulushlari va xato turlari bo'yicha chalkashish matritsasini berish kerak. Biror sinf uchun musbat kuzatuv juda oz bo'lsa, bitta intervaldagi qiymatni qat'iy xulosa sifatida sharhlash o'rniga noaniqlikni ko'rsatish maqsadga muvofiq.

Pilot bosqich, ikki annotatorli sxema va validatsiya protokoli

Asosiy annotatsiya uchun minimal tartib ikki mustaqil yozuv (A va B), umumiy scheme_version, bir xil source_version va ikki annotatorga berilgan kontekst chegaralarini nazarda tutadi. Mustaqil belgilash tugagach, xom yozuvlar muzlatiladi; faqat shundan keyin mavjudlik, span va kategoriya kelishuvi hisoblanadi, so‘ng kelishmovchiliklar sabablar bo‘yicha ajratilib ekspert hal qilishi uchun yuboriladi.

Barcha belgilarni teng hajmda ikki marta tekshirish vaqt jihatidan qimmat bo‘lganda, jarayon bosqichlarga ajratiladi: tanlangan, qatlamlangan qism to‘liq ikki mustaqil annotatsiyadan o‘tadi, qolgan qismda esa birlamchi belgilash va riskka asoslangan ekspert nazorati qo‘llanadi. Bunday holda faqat tanlab ikki marta belgilangan qatlam bo‘yicha hisoblangan kelishuv umumiy bir annotatorli qismga avtomatik ko‘chirilmaydi.

Annotatsiya sifatining texnik tomoni lingvistik ekspertizadan keyin emas, jarayon davomida tekshiriladi: validator har bir span chegarasi matn uzunligi doirasida ekanini, interval original_fragment bilan mosligini, kategoriya yorlig‘i ruxsat etilgan sxemaga tegishli ekanini va manba versiyasi o‘zgarmaganini nazorat qiladi.

2-jadval. Annotatsiyalash jarayonining pilot bosqichlari, tekshiriladigan sifat belgilari va shakllantiriladigan artefaktlar

Bosqich	Tekshiriladigan sifat belgisi	Chiqariladigan artefakt
Dizayn	Teg va istisnolar ta’rifi	Annotatsiyalash yo‘riqnomasi v1
Pilot	Noaniq teglar, span kelishmovchiligi	Yangi yo‘riqnoma va nizo ro‘yxati
Mustaqil belgilash	A/B yozuvlarining mustaqilligi	raw_A, raw_B, versiya xeshi
Statistik audit	Mavjudlik, chegara, kategoriya	κ , α , span diagnostikasi protokoli
Adjudikatsiya	Dalilli ekspert qarori	gold va qaror jurnali
Texnik validator	Ofset, ID, ruxsat etilgan qiymat	Xatolar va tekshiruv jurnali
Resurs pasporti	Janr/soha/huquqiy maqom	Dataset pasporti va cheklovlar



2-rasm. Ishonchlilikni baholash jarayonining olti bosqichi

Xulosa. Ko'p qatlamli annotatsiya sxemasining ilmiy qiymati faqat teglar tasnifi bilan emas, balki uning ishonchliligini qanday o'lchash mumkinligi bilan ham belgilanadi. Maqolada ko'rsatilganidek, bitta umumiy kelishuv foizi hodisaning mavjudligi, span chegarasi va kategoriya tanlovidagi farqlarni bir-biridan ajrata olmaydi – shuning uchun har bir komponent uchun mos o'lchov (Cohen's κ , Krippendorff's α , γ yoki exact-match F_1) alohida qo'llanilishi zarur. Koeffitsiyentlarni talqin qilishda yagona universal chegaradan voz kechish, kamyob sinflar va “prevalence paradox” kabi xavflarni hisobga olish, shuningdek ikki annotatorli pilot sxema va deterministik texnik validatorga tayanish – bularning barchasi oltin standart annotatsiyaning ilmiy shaffofligini ta'minlaydigan shartlardir.

Taklif etilgan baholash protokoli avtomatik til modellari uchun ishonchli o'quv va sinov ma'lumotlarini shakllantirishning zaruriy bosqichi bo'lib, uning amaliy natijalari – hisoblangan koeffitsiyentlar va annotatorlar kelishuvi – real korpusda o'tkaziladigan keyingi empirik tadqiqot doirasida taqdim etiladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Pustejovsky J., Stubbs A. Natural Language Annotation for Machine Learning. Sebastopol: O'Reilly Media; 2012.
2. Artstein R., Poesio M. Survey Article: Inter-Coder Agreement for Computational Linguistics. Computational Linguistics. 2008;34(4):555–596. doi:10.1162/coli.07-034-R2.

3. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960;20(1):37–46. doi:10.1177/001316446002000104.
4. Krippendorff K. Reliability in Content Analysis: Some Common Misconceptions and Recommendations. Human Communication Research. 2004;30(3):411–433. doi:10.1111/j.1468-2958.2004.tb00738.x.
5. Mathet Y., Widlöcher A., Métivier J.-P. The Unified and Holistic Method Gamma (γ) for Inter-Annotator Agreement Measure and Alignment. Computational Linguistics. 2015;41(3):437–479. doi:10.1162/COLI_a_00227.
6. Bayerl P. S., Paul K. I. What Determines Inter-Coder Agreement in Manual Annotations? A Meta-Analytic Investigation. Computational Linguistics. 2011;37(4):699–725. doi:10.1162/COLI_a_00074.
7. Lombard M., Snyder-Duch J., Bracken C. C. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. Human Communication Research. 2002;28(4):587–604. doi:10.1111/j.1468-2958.2002.tb00826.x.
8. Krippendorff K. Content Analysis: An Introduction to Its Methodology. 4th ed. Los Angeles: SAGE; 2018.