1 Introduction

What's the difference between task-oriented and non-task-oriented systems?

Task-oriented systémy se zaměřují na splnění konkrétního úkolu nebo sady úkolů, například rezervaci restaurace či letu, a často využívají úzce specializované backendové rozhraní. Non-task-oriented systémy naproti tomu slouží k otevřené, volné konverzaci za účelem zábavy či budování společenské interakce, aniž by jejich dialog směřoval k dosažení specifického cíle. Zatímco task-oriented systémy kladou důraz na efektivitu dosažení cíle, non-task-oriented systémy se soustředí na plynulost a přirozenost rozhovoru.

Describe the difference between closed-domain, multi-domain, and open-domain systems.

Closed-domain systémy operují v úzce vymezené oblasti s omezeným souborem témat a úkolů, což umožňuje vysokou přesnost v dané doméně. Multi-domain systémy spojují několik takových uzavřených domén a dokážou se přepínat mezi různými oblastmi podle požadavků uživatele. Open-domain systémy (často založené na LLM) naopak podporují volný rozhovor na širokou škálu témat bez předem definovaných omezení, ale s nižší specializací a potenciálně vyšší chybovostí.

Describe the difference between user-initiative, mixed-initiative, and system-initiative systems.

System-initiative systémy řídí tok dialogu tím, že systém klade otázky a uživatel na ně postupně odpovídá, což zajišťuje robustnost, ale může působit neflexibilně. User-initiative systémy umožňují uživateli převzít iniciativu a definovat, čeho chce dosáhnout, přičemž systém se pouze přizpůsobuje jeho požadavkům (např. Nastav budík na dvě minuty“). Mixed-initiative systémy pak kombinují obě strategie, kdy jak uživatel, tak systém mohou klást otázky a navrhovat kroky, což vede k přirozenějšímu, avšak složitějšímu dialogu.

2 Linguistics of Dialogue

What are turn taking cues/hints in a dialogue? Name a few examples.

V dialogu slouží signály pro řízení předávání slova k označení konce či začátku turnu. Patří mezi ně lingvistické ukazatele, například dokončené věty nebo specifické fráze, prosodické znaky jako změna výšky hlasu či délka pauzy mezi dvěma projevy, a neverbální projevy – oční kontakt, gesta nebo mimika. Tyto náznaky pomáhají minimalizovat překryvy a ticha a udržet plynulost konverzace

Explain the main idea of the speech acts theory.

Teorie řečových aktů rozlišuje úroveň samotného pronesení slov (utterance act), jejich doslovný význam (propositional act), záměr mluvčího (illocutionary act) a skutečný dopad na posluchače (perlocutionary act). Hlavní myšlenkou je, že každým výroky nejen sdělujeme informace, ale i provádíme určité činnosti

– žádáme, přikazujeme, slibujeme či vyjadřujeme pocity. Tím se jazyk stává nástrojem pro akci ve světě, nikoli pouze popisem reality

What is grounding in dialogue?

Grounding je proces zajišťující vzájemné porozumění mluvčího a posluchače během konverzace. Vychází z předpokladu sdílené znalosti (common ground), kterou účastníci postupně rozšiřují a ověřují vzájemnými signály. Cílem je potvrdit, že sdělení bylo správně přijato a pochopeno, než se pokračuje dál

Give some examples of grounding signals in dialogue.

Mezi pozitivní signály porozumění patří zpětné kanály (uh-huh“, ano“), úsměv či přikývnutí, explicitní potvrzení (Rozumím“) i implicitní pokračování tématu. Negativní signály zahrnují zaskočené ticho, výrazy pochybnosti nebo žádosti o opravu (Promiň, co tím myslíš?“). Dále sem patří repair requests a clarification requests, které ukazují, že je třeba upřesnit či opravit předchozí výpověď

What is deixis? Give some examples of deictic expressions.

Deixis označuje jazykové prostředky ukazující na kontextové prvky jako prostor, čas či osoby. Mezi deiktika patří osobní zájmena (já“, ty“), ukazovací zájmena (tento“, tamten“), časová označení (nyní“, včera“) či místní příslovce (zde“, tam“). Význam těchto výrazů závisí na situaci, ve které se rozhovor odehrává, a na společném kontextu mluvčích

What is coreference and how is it used in dialogue?

Koreference (anaphora) spočívá v tom, že jeden výraz odkazuje zpětně na dříve zmíněný objekt či entitu (např. Marie… ona“). Slouží k tomu, aby se předešlo opakování plných názvů a aby text či řeč zůstaly kompaktní a srozumitelné. V dialogových systémech se koreference využívá k udržení kontextu, například při sledování, na co uživatel v předchozí větě odkazoval

What does Shannon entropy and conditional entropy measure? No need to give the formula, just the principle.

Shannonova entropie měří průměrné množství nejistoty nebo překvapení“ ve zprávě, tedy jak moc informací přináší každý prvek vzhledem k pravděpodobnostnímu rozdělení. Podmíněná entropie hodnotí, jaká je zbývající nejistota o dalším prvku (např. slově) poté, co známe kontext (např. předchozí slova). Tyto principy se v dialogových systémech využívají při modelování jazyků a předpovídání následujícího slova či fráze

What is entrainment/adaptation/alignment in dialogue?

Entrainment (také alignment či adaptace) je jev, kdy si partneři v konverzaci postupně přizpůsobují řečový styl, volbu slov, syntaktické konstrukce nebo prosodii. Tento jev podporuje plynulost a přirozenost dialogu a pomáhá budovat důvěru mezi mluvčími. V systémech to znamená, že by se stroj měl adaptovat na uživatelův jazyk, aby bylo porozumění efektivnější a interakce přirozenější

3 Data & Evaluation

What are the typical options for collecting dialogue data?

Mezi běžné způsoby sběru dialogových dat patří interní sběr v laboratoři (in-house), kde experti nebo studenti vedou a anotují rozhovory, což je kvalitní, ale nákladné a časově náročné. Další možností je automatické prohledávání webu (web crawling), které je rychlé a levné, ale často obsahuje nerelevantní či nekvalitní konverzace. Častým kompromisem je crowdsourcing, kdy se úkoly zadávají masám online pracovníků, což umožňuje rychlý sběr velkého množství dat, ale kvalita může být proměnlivá.

How does Wizard-of-Oz data collection work?

Metoda Wizard-of-Oz spočívá v tom, že účastníci věří, že komunikují s automatizovaným systémem, zatímco ve skutečnosti jejich interakce ručně řídí člověk kouzelník". Kouzelník za scénou přijímá vstupy uživatele a v reálném čase generuje odpovědi buď výběrem z předdefinovaných možností, nebo volným psaním. Tento přístup umožňuje simulovat funkčnost systému, který ještě není implementován, a sbírat přitom autentické reakce uživatelů.

What is corpus annotation, what is inter-annotator agreement?

Anotace korpusu znamená přidání štítků nebo metadat k nasbíraným dialogům, např. přepis mluvené řeči, klasifikaci dialogových aktů či označení entit. Inter-annotator agreement (IAA) je míra, jak moc se různí anotátoři shodují na přiřazených štítcích, často se počítá pomocí Cohenova kappa koeficientu. Vysoká hodnota IAA zajišťuje, že anotace jsou spolehlivé a interpretace dat není příliš subjektivní.

What is the difference between intrinsic and extrinsic evaluation?

Intrinsic evaluation hodnotí vlastnosti komponenty nebo modelu izolovaně, například kvalitu výstupu NLG měřenou n-gramovou podobností. Extrinsic evaluation měří dopad daného systému v reálném kontextu použití, tedy jak dobře plní skutečný úkol nebo jaké přináší uživatelské zkušenosti. Zatímco intrinsic testy zkoumají interní vlastnosti, extrinsic testy ověřují praktickou užitečnost.

What is the difference between subjective and objective evaluation?

Objective evaluation používá automatické metriky založené na datech, například WER pro ASR nebo F1-skóre pro NLU, a je plně reprodukovatelná. Subjective evaluation se opírá o lidské hodnocení, typicky formou dotazníků nebo intervencí, kde uživatelé hodnotí spokojenost, přirozenost či srozumitelnost. Objektivní metriky jsou rychlé a levné, ale nemusí vždy korelovat s lidskou percepcí, zatímco subjektivní testy zohledňují skutečné vnímání.

What are the main extrinsic evaluation techniques for task-oriented dialogue systems?

Pro task-oriented systémy se jako objektivní extrinsic metriky používá míra úspěšnosti úkolu (goal completion rate), počet dialogových kol nebo doba konverzace (méně kol či kratší doba značí efektivitu) a poměr selhání (fallback rate). Subjektivně se hodnotí pomocí uživatelských dotazníků s Likertovou škálou na otázky typu Dostal jsem požadované informace" či Chtěl(a) bych systém znovu použít". Kombinace obou přístupů poskytuje nejkomplexnější obraz o výkonu v reálných podmínkách.

What are some evaluation metrics for non-task-oriented systems (chatbots)?

U chitchat botů se objektivně často měří délka konverzace (počet výměn či čas), přičemž delší interakce obecně znamená vyšší zapojení uživatele. Dále lze sledovat míru návratu uživatelů (returning users) nebo rozmanitost odpovědí (distinct-1/2 – podíl jedinečných n-gramů). Subjektivně se hodnotí zábavnost, přirozenost a spokojenost, obvykle dotazníky typu Bavilo mě konverzovat“ či Byl chatbot sympatický“.

What's the main metric for evaluating ASR systems?

Hlavní metrika pro hodnocení ASR systémů je Word Error Rate (WER), která vyjadřuje poměr chyb (substituce, inserce, deleti) k celkové délce referenčního textu. Nižší WER znamená, že rozpoznávání je přesnější a odpovídá zápisu lidského transkriptu. WER umožňuje srovnání různých ASR řešení na jednotné stupnici.

What's the main metric for NLU (both slots and intents)?

Pro vyhodnocení slotů v NLU se používá F1-skóre, které kombinuje přesnost (precision) a úplnost (recall) vyextrahovaných entit. Pro klasifikaci záměrů (intents) se nejčastěji měří přesnost (accuracy), tedy podíl správně rozpoznaných úmyslů vzhledem k referenční anotaci. Tyto metriky dávají přehled o tom, jak spolehlivě model rozumí struktuře a obsahu uživatelských požadavků.

Explain an NLG evaluation metric of your choice.

BLEU je metrika pro hodnocení NLG založená na podobnosti n-gramů generovaného textu vůči jednomu či více referenčním překladům. Zahrnuje přesnost pro různé délky n-gramů a penalizuje příliš krátké výstupy brevity penalty. Využívá se zejména u strojového překladu a někdy u NLG, i když její korelace s lidským hodnocením na úrovni jednotlivých vět je omezená.

Why do you need to check for statistical significance (when evaluating an NLP experiment and comparing systems)?

Statistickou významnost je nutné ověřit, aby bylo možné prokázat, že rozdíl ve výkonech není dílem náhody, ale skutečným zlepšením modelu. Testy jako Studentův t-test nebo bootstrap resampling umožňují stanovit, zda lze nulovou hypotézu (žádný rozdíl“) zamítnout s dostatečnou důvěrou (např. 95 %). Bez těchto testů hrozí, že na základě náhodných fluktuací dat budeme chybně hodnotit lepší či horší chování systému.

Why do you need to evaluate on a separate test set?

Samostatný testovací soubor je potřeba k objektivnímu posouzení schopnosti modelu generalizovat na dosud neviděná data a zabránit overfittingu. Pokud by se systém ladil i na testovací data, mohlo by dojít k učení se zpaměti“ konkrétních příkladů, což zkreslí reálnou výkonnost. Oddělením tréninkové, validační a testovací množiny se zajistí, že výsledky odrážejí skutečnou robustnost a použitelnost modelu.

4 Natural Language Understanding

What are some alternative semantic representations of utterances, in addition to dialogue acts?

Kromě dialogových aktů se v NLU používají i hierarchické syntaktické či sémantické stromy, rámcové struktury (frames) se sloty a podrámci nebo grafové reprezentace jako AMR. Stromové a rámcové formy zachycují vztahy mezi jednotlivými složkami věty, zatímco grafy umožňují modelovat složitější závislosti a sdílené podvýrazy. Tyto alternativy se hodí pro hlubší sémantické parsování a generalizaci mimo úzkou doménu.

Describe language understanding as classification and language understanding as sequence tagging.

Při přístupu jako klasifikace se každé úmyslové nebo slotové koncepty (intent/slot-value) detekují nezávislými binárními klasifikátory, které rozhodují, zda se daný koncept v uttenance vyskytuje. Přístup jako označování sekvence (sequence tagging) přiřazuje ke každému slovu v větě štítek (např. IOB formát), čímž lze přímo extrahovat rozsahy slotů. Zatímco klasifikace je jednodušší a nezávislá, sekvenční označování umožňuje lepší zachycení kontextu a hierarchie uvnitř věty.

How do you deal with conflicting slots or intents in classification-based NLU?

Konflikty mezi sloty nebo úmysly se obvykle řeší externím mechanismem, který porovnává konfigurační skóre (confidence) jednotlivých klasifikátorů a vybírá nejpravděpodobnější kombinaci. Může se také aplikovat sada pravidel, která v případě neslučitelnosti upřednostní některé sloty nebo zamítne celý turn pro explicitní dotaz na upřesnění. V některých systémech se konflikty detekují jako chyby a spustí se proces objasňování (clarification).

What is delexicalization and why is it helpful in NLU?

Delexikalizace znamená nahrazení konkrétních hodnot slotů (např. jmena měst, restaurací) zástupnými symboly `<slot>` v trénovacích větách. Tím se výrazně snižuje sparsita dat a model se lépe generalizuje na neznámé hodnoty, protože se místo tisíců možných entit učí vzory okolo placeholderů. Delexikalizace také umožňuje snadnější přenos modelu do jiných domén jen změnou slovníku zástupných symbolů.

Describe one of the approaches to slot tagging as sequence tagging.

Jednou z metod je použití lineárně-řetězcového CRF, který modeluje pravděpodobnost celé sekvence IOB štítků vzhledem k pozorované větě a zajišťuje globální normalizaci. CRF bere v úvahu nejen vlastnosti jednotlivých slov, ale i přechody mezi sousedními štítky, což zabraňuje neplatným kombinačním sekvencím. Díky tomu jsou výsledky konzistentnější než u lokálně normalizovaných modelů jako MEMM.

What is the IOB/BIO format for slot tagging?

IOB (také nazývaný BIO) formát přiřazuje ke každému tokenu jeden ze štítků: `B-<slot>` pro začátek hodnoty slotu, `I-<slot>` pro pokračování stejného slotu a `O` pro tokeny mimo slot. Tento formát umožňuje jednoznačně znázornit

hranice víceslovných entit a snadno rekonstruovat jejich rozsah. Díky IOB lze rozlišit sousední entitu stejného typu a zabránit jejich slévání.

What is the label bias problem?

Label bias je zkreslení lokálně normalizovaných sekvenčních modelů (např. MEMM), kde stavy s méně možnými přechody zdánlivě získávají nepřiměřeně vysokou pravděpodobnost. Model tak upřednostňuje přechody z těchto úzkých" stavů bez ohledu na celkový kontext, což vede k chybám ve vyhledávání optimální sekvence. Tento problém řeší globálně normalizované modely jako CRF.

How can an NLU system deal with noisy ASR output? Propose an example solution.

Pro zvýšení odolnosti vůči chybám ASR lze místo jediného přepisu zpracovávat n-best seznam hypotéz nebo konfuzní síť, kde se berou v úvahu alternativní výsledky rozpoznání se svou váhou". NLU pak vyhodnocuje sloty a intenty přes všechny varianty, případně agreguje skóre napříč hypotézami a vybírá nejpravděpodobnější s ohledem na ASR confidence. Tímto způsobem se snižuje vliv jednotlivých chyb a zvyšuje celková robustnost rozpoznání záměrů.

5 Neural NLU & Dialogue State Tracking

Describe a neural architecture for NLU.

Typická neuronová architektura pro NLU začíná vrstvou vstupních embeddingů slov, které reprezentují každý token jako vektor. Následuje sekvenční encoder, nejčastěji bidirekcionální LSTM/GRU nebo Transformer, který zpracuje kontext okolních slov. Na výstup se aplikuje buď softmax pro klasifikaci intentu a slotů (případně CRF pro sekvenční štítkování), nebo dekodér s attention pro generování štítků po jednotlivých krocích.

What is the dialogue state and what does it contain?

Dialogový stav je strukturovaný souhrn informací, které systém dosud získal, a obsahuje preference uživatele (sloty a jejich hodnoty), položky, o něž uživatel požádal, již potvrzené nebo odmítnuté sloty, seznam navržených možností a historii systémových akcí. Dále může zahrnovat metadata jako potvrzení, žádosti o opravu či klíčové signály konverzační situace (např. restart" či nashledanou"). Tento stav slouží jako vstup pro rozhodování dialogového manažera.

What is an ontology in task-oriented dialogue systems?

Ontologie v task-oriented systémech definuje slovník všech slotů a jejich povolených hodnot, stejně jako kategorie akcí (inform, request) a závislosti mezi koncepty. Umožňuje systémům vědět, které sloty jsou informovatelné nebo dotazovatelné a jaké hodnoty lze pro daný slot akceptovat. Ontologie tak vytváří rámec, podle něhož mohou NLU komponenty a tracker porozumět a ověřit konzistenci uživatelových požadavků.

Describe the task of a dialogue state tracker.

Úkolem trackeru dialogového stavu je na základě nového uživatelova vstupu (NLU výstupu) aktualizovat interní reprezentaci stavu tak, aby odrážela nejpravděpodobnější preference a požadavky uživatele. To zahrnuje začlenění nově identifikovaných slotů, potvrzení či opravy předchozích hodnot a zachování

historie konverzace. Výsledek pak slouží dialogovému manažeru pro výběr další systémové akce.

What's a partially observable Markov decision process?

Partially Observable Markov Decision Process (POMDP) je rozšíření MDP, kde není možné přímo pozorovat skutečný stav procesu, ale dostáváme jen pravděpodobnostní pozorování. Systém proto udržuje belief state – rozložení pravděpodobností přes možné skryté stavy – a pomocí bayesovské aktualizace ho při každém kroku upravuje podle akce a nového pozorování. POMDP modeluje dialog jako náhodný proces se stavy, akcemi, pozorováními a odměnami.

Describe a viable architecture for a belief state tracker.

Jedna životaschopná architektura pro sledování stavu (belief tracking) využívá rekurentní neuronové sítě, například LSTM. Síť postupně zpracovává historii dialogu (uživatelské i systémové repliky) a vytváří vnitřní reprezentaci kontextu. Pro každý slot pak navazuje klasifikátor (např. softmax nebo predikce rozsahu), který určuje pravděpodobné hodnoty slotu — tento přístup umožňuje průběžné sledování změn a zvládá i neurčitost ve vstupu.

What is the difference between dialogue state and belief state?

Dialogový stav je jednorázová, deterministická reprezentace (1-best) obsahu dialogu z předchozích kol (sloty, hodnoty, akce). Belief state je naopak pravděpodobnostní distribuce přes možné stavy, uchovávající nejistotu a kumulující důvěryhodnost opakovaných nebo protichůdných vstupů. Zatímco dialogový stav může být chybný při nejistých NLU výstupech, belief state zpracovává všechny hypotézy konzistentně.

What's the difference between a static and a dynamic state tracker?

Statický tracker kódováním historie vytváří pevný vektor rysů z předchozích otázek a odpovědí (např. součty, sliding window) a jednorázovým klasifikátorem odhaduje stav. Dynamický tracker naopak modeluje dialog jako sekvenci a využívá sekvenčních modelů (CRF, RNN) pro explicitní učení závislostí a přechodů mezi stavy napříč koly. Dynamické přístupy tak lépe zachycují tok konverzace a kontextuální vlivy.

How can you use pretrained language models in NLU?

Předtrénované jazykové modely lze využít pro rozpoznávání záměrů (intents), klasifikaci vět a extrakci entit. Díky svému širokému jazykovému pokrytí a znalostem mohou poskytovat velmi dobré výsledky i bez rozsáhlého jemného doladění. Mohou být použity jako embedding vrstva nebo jako celý klasifikátor s doladěním na konkrétní úlohu.

How can you use pretrained language models or large language models for state tracking?

Velké jazykové modely mohou sledovat stav dialogu implicitně, tedy bez nutnosti explicitní reprezentace. Na základě celé historie konverzace dokážou odpovědět nebo navrhnout další krok, aniž by si uchovávaly strukturovaný stav. Alternativně je lze využít i pro predikci slotů či formulaci aktualizovaného stavu pomocí generativního přístupu.

6 Dialogue Policies

What are the non-statistical approaches to dialogue management/action selection?

Mezi nestatistické přístupy patří konečné automaty (FSM), kde je dialog předem definován jako posloupnost stavů a přechodů. Dále se používají rámcové (frame-based) systémy, které podle předem určené sady slotů dynamicky doplňují potřebné informace v libovolném pořadí. K tomu se často přidávají pravidlové (rule-based) metody, kdy akce vybírají ručně napsaná if-then-else pravidla nad věrohodnostmi jednotlivých slotů.

Why is reinforcement learning preferred over supervised learning for training dialogue managers?

Reinforcement learning umožňuje učit se strategii přímo ze zpětné vazby (reward), aniž bychom potřebovali rozsáhlá anotovaná data se správnými akcemi. Dialogy jsou mnohoznačné a v reálných situacích může být několik stejně dobrých odpovědí, takže naučit se pouze z jednoho správného“ příkladu by vedlo k omezené robustnosti. RL navíc podporuje exploraci nových dialogových cest, na které bychom v tréninkových datech nemuseli narazit.

Describe the main idea of reinforcement learning (agent, environment, states, rewards).

V RL agenta představuje dialogový manažer, který v každém kroku pozoruje stav prostředí (dialogový stav) a provede akci (např. systémový dialogový akt). Prostředí mu na základě této akce vrátí odměnu (reward) a nový stav, což umožňuje odhadnout, jak dobré je dané chování z dlouhodobého hlediska. Cílem je najít politiku (mapování stav→akce), která maximalizuje kumulativní očekávanou odměnu.

What are deterministic and stochastic policies in dialogue management?

Deterministická politika vždy ve stejném stavu zvolí tu samou akci, což odpovídá přímému pravidlu nebo tabulce stav–akce. Stochastická politika vrací pro každý stav distribuci pravděpodobností nad akcemi a konkrétní akce se pak vybere náhodným vzorkováním. Stochastické přístupy podporují exploraci a zabraňují zasekání se v lokálních minimech.

What's a value function in a reinforcement learning scenario?

Value funkce $V^\pi(s)$ udává očekávanou kumulativní odměnu, kterou agent získá, když začne ve stavu s a bude následovat politiku π. Slouží k porovnání, které stavy jsou výhodnější z dlouhodobého hlediska. Podobně Q-funkce $Q^\pi(s, a)$ rozšiřuje hodnotu o konkrétní počáteční akci a.

What's the difference between actor and critic methods in reinforcement learning?

Actor-kritik architektury kombinují dvě komponenty: actor (politik) rozhoduje o tom, jakou akci zvolit, a critic (hodnotitel) odhaduje hodnotu stavů nebo stav-akce (value/Q funkci). Critic poskytuje actorovi zpětnou vazbu ve formě gradientů, čímž zlepšuje učení politiky. Díky tomu dohromady dosahují stabilnějšího a rychlejšího konvergence než čistě policy-gradient nebo čistě value-based metody.

What's the difference between model-based and model-free approaches in RL?

Model-based metody pracují s odhadnutým nebo známým modelem přechodových pravděpodobností a odměn $p(s'|s,a)$ a $r(s,a,s')$, což umožňuje plánování (např. pomocí dynamic programming). Model-free přístupy se učí přímo z nasbíraných zkušeností bez předpokladu známého modelu, např. Q-learning nebo policy gradient. Model-free metody jsou jednodušší na nasazení v reálném prostředí, kde je model obvykle neznámý.

What are the main optimization approaches in reinforcement learning (what measures can you optimize and how)?

V RL lze optimalizovat hodnotové funkce (value-based) jako $V(s)$ nebo $Q(s,a)$ pomocí algoritmů jako SARSA, Q-learning nebo TD-učení, které se opírají o Bellmanovy aktualizace. Alternativně lze přímo optimalizovat politiku (policy-based) pomocí policy gradient metod typu REINFORCE nebo actor-critic. Dalšími přístupy jsou model-based optimalizace s dynamic programming nebo Monte Carlo metody pro odhad vrácených odměn.

Why do you typically need a user simulator to train a reinforcement learning dialogue policy?

RL vyžaduje obrovské množství interakcí, často tisíce až statisíce dialogů, což není reálně udržitelné se skutečnými uživateli. Simulátor uživatele napodobuje chování člověka na úrovni dialogových aktů a umožňuje efektivní on-policy učení a exploraci bez nutnosti živých testů. Díky simulátoru lze navíc dobře kontrolovat hladinu šumu a chyby ASR/NLU, což zrychluje vývoj a ladění politiky.

7 Neural Policies & Natural Language Generation

How do you involve neural networks in reinforcement learning (describe a Q network or a policy network)?

Neurony se v RL používají jako aproximátor hodnotové funkce nebo politiky – v Deep Q-Network je Q-funkce $Q(s,a;\theta)$ modelována sítí, která přijímá stav jako vstup a vypočítá hodnoty pro jednotlivé akce. V policy-gradient metodách (REINFORCE či actor-critic) je politika $\pi(a|s;\theta)$ parametrizována sítí, která přímo produkuje pravděpodobnosti nebo logity akcí. Parametry sítě se učí gradientním sestupem na základě odhadů návratnosti či odměnových signálů při interakci s prostředím.

What are the main steps of a traditional NLG pipeline – describe at least 2.

Tradiční NLG pipeline začíná *content planning*, kde se vybírá a strukturuje, jaké informace (fakta) se mají v textu objevit. *Sentence planning* (mikroplánování) pak provádí agregaci faktů do vět, volbu slov a referenčních výrazů. V poslední fázi *surface realization* se tyto plány převedou podle gramatických pravidel v lineární text s korektní morfologií a slovosledem.

Describe one approach to NLG of your choice.

Template-based NLG používá předdefinované šablony obsahující zástupné místo pro konkrétní hodnoty slotů, které se do nich doplní. Šablony se obvykle navrhují ručně pro každý dialogový akt a pokrývají nejčastější scénáře – systém pak podle aktů vybere odpovídající vzor a vyplní jeho parametry. Tento přístup je spolehlivý a umožňuje do něj zapracovat pravidla pro skloňování či volbu členu.

Describe how template-based NLG works.

V template-based NLG se definují textové vzory jako řetězce s proměnnými (např. `"Restaurant {name} is located at {address}."`), které odpovídají jednotlivým systémovým akcím. Když přijde dialogový akt se sloty, engine vybere vhodnou šablonu podle typu a do vyznačených míst dosadí konkrétní hodnoty. Výsledný text se pak případně prožene jednoduchými pravidly pro inflexi a volbu správných tvarů.

What are some problems you need to deal with in template-based NLG?

V template-based přístupu je nutné řešit správné skloňování a shodu podmětu s přísudkem při různých vložených hodnotách a generovat korektní tvary slov. Šablony postrádají variabilitu řeči, takže je třeba ručně vytvářet více variant, aby odpovědi nepůsobily jednotvárně. Dále nelze snadno škálovat na nové domény bez rozsáhlé ruční práce a přidávání nových slotů vyžaduje úpravu všech souvisejících šablon.

Describe a possible neural networks based NLG architecture.

Jednoduchá architektura pro generování textu pomocí neuronových sítí je model encoder-decoder. Enkodér převede vstupní informaci (např. seznam slotů nebo dialogový akt) na vnitřní vektorovou reprezentaci. Dekodér pak generuje výstupní větu po slovech – jedno slovo za druhým. Pomocí attention se při generování každého slova soustředí na relevantní části vstupu. Modernější verze používá Transformer architekturu, která díky samoattention zvládne delší vstupy i výstupy najednou a bez nutnosti krokovat sekvenci po jednom.

How can you use pretrained language models or large language models in NLG?

V NLG slouží velké jazykové modely jako generátory přirozeného textu na základě strukturovaného vstupu, např. slotů nebo klíčových bodů. Dokážou produkovat plynulé, srozumitelné a kontextově vhodné věty. Lze je použít buď přímo (promptingem) nebo s jemným doladěním na konkrétní výstupový styl či doménu.

8 Voice Assistants & Question Answering

What is a smart speaker made of and how does it work?

Chytrý reproduktor obsahuje mikrofony (často více pro zpracování vzdáleného zvuku), reproduktor, procesor (CPU), paměť, Wi-Fi a Bluetooth modul. Po aktivaci budicí frází (Hey Siri“, Alexa“) se zaznamenaný zvuk odešle do cloudu, kde proběhne rozpoznání řeči, pochopení záměru a případně získání odpovědi.

Výsledek je pak přehrán uživateli nebo provede požadovanou akci (např. přehrání hudby, nastavení budíku).

Briefly describe a viable approach to question answering.

Základní přístup k QA kombinuje vyhledání relevantních dokumentů nebo pasáží a následné extrahování odpovědi. Nejprve se použije IR metoda (např. TF-IDF, BM25 nebo dense retrieval) k nalezení dokumentů, poté se z nich pomocí pravidel nebo neuronových sítí extrahuje konkrétní entita nebo fráze. Případně se z těchto pasáží vygeneruje odpověď pomocí jazykového modelu.

What is document retrieval and how is it used in question answering?

Document retrieval je proces, kdy se k dotazu vyhledají nejrelevantnější dokumenty z velkého korpusu pomocí metrik jako TF-IDF nebo BM25. V QA systému tento krok zužuje prostor, ze kterého se odpověď extrahuje, čímž šetří výpočetní čas a zvyšuje přesnost. Dokumenty jsou následně použity jako vstup pro extrakci nebo generování odpovědi.

What is dense retrieval?

Dense retrieval používá neuronové modely k zakódování dotazu i dokumentů do vektorového prostoru, kde se hledají nejbližší shody podle kosinové podobnosti. Vektory dokumentů se předpočítají a uloží, což umožňuje rychlé vyhledávání i ve velkých kolekcích. Tento přístup zachycuje sémantickou podobnost lépe než tradiční metody založené na klíčových slovech.

How can you use neural models in answer extraction?

Neuronové modely, jako BERT, se použijí tak, že jako vstup dostanou otázku a pasáž textu, ze které mají najít odpověď. Model predikuje pravděpodobnosti začátku a konce odpovědi v textu pomocí softmax klasifikace nad jednotlivými tokeny. Výstupem je konkrétní úsek textu (span), který s největší pravděpodobností obsahuje odpověď.

How can you use retrieval-augmented generation in question answering?

Retrieval-augmented generation (RAG) kombinuje vyhledávání a generování: nejprve se najdou relevantní texty a poté se použijí jako vstup pro jazykový model (např. T5 nebo GPT), který na jejich základě vygeneruje odpověď. Tato metoda umožňuje odpovídat i na otázky vyžadující kombinaci více informací a zároveň snižuje riziko halucinací. Retriever i generátor mohou být trénovány společně.

What is a knowledge graph?

Knowledge graph (KG) je strukturovaná databáze znalostí, kde jsou informace reprezentovány jako trojice ⟨subjekt, predikát, objekt⟩. Obsahuje entity (např. lidé, místa, objekty) a vztahy mezi nimi, přičemž každá entita i vztah má typ a může být součástí ontologie. KG umožňuje přesné a efektivní odpovídání na faktografické dotazy, často pomocí dotazovacího jazyka SPARQL.

9 Dialogue Tooling

What is a dialogue flow/tree?

Dialogový tok (flow) je struktura, která definuje průběh konverzace. Může být lineární (slot-filling), kde se postupně sbírají informace, nebo stromová, kde přechody mezi uzly závisí na podmínkách a volbách uživatele. Umožňuje flexibilní reakce a návraty k předchozím bodům.

What are intents and entities/slots?

Intents (záměry) popisují, co chce uživatel udělat (např. rezervace stolu). Entities nebo slots jsou konkrétní informace potřebné k realizaci daného záměru (např. čas, počet lidí). Rozpoznávání těchto prvků je klíčové pro porozumění uživatelskému vstupu.

How can you improve a chatbot in production?

Chatbot lze zlepšovat analýzou logů a reakcí uživatelů – například sledováním, které odpovědi byly vybrány. Lze také měřit metriky jako coverage (pokrytí) a containment (úspěšnost bez zásahu člověka). Na základě těchto dat lze chatbot aktualizovat a lépe přizpůsobit uživatelům.

What is the containment rate (in the context of using dialogue systems in call centers)?

Containment rate je podíl konverzací, které chatbot zvládl bez zásahu člověka. Ukazuje, jak efektivně dokáže chatbot vyřešit požadavky uživatelů. Vysoké skóre znamená méně přepojování na operátory a vyšší efektivitu.

What is retrieval-augmented generation?

Retrieval-augmented generation (RAG) je technika, která kombinuje vyhledávání dokumentů s generováním odpovědi. Nejprve se najdou relevantní texty (retrieval), a následně se použijí jako vstup pro generativní model, který vytvoří odpověď. Výsledkem je informovaná a přirozená odpověď.

10 Automatic Speech Recognition

What is a speech activity detector?

Detektor řečové aktivity (VAD) je komponenta, která rozlišuje mezi řečí a ostatními zvuky. Používá se k tomu, aby se systém spustil jen tehdy, když někdo mluví, čímž šetří výpočetní výkon a snižuje chybovost. Moderní VAD využívá statistické nebo neuronové modely a funguje jako binární klasifikátor.

Describe the main components of an ASR pipeline system.

Typická ASR pipeline se skládá z detektoru řeči (VAD), extrakce featur, akustického modelu, jazykového modelu a dekodéru. Akustický model předpovídá fonémy ze zvukových vlastností, jazykový model určuje pravděpodobnost posloupnosti slov a dekodér vybírá nejpravděpodobnější text. Každá část zpracovává jiný aspekt vstupního audia.

How do input features for an ASR model look like?

Vstupními featurami jsou obvykle Mel-frekvenční cepstrální koeficienty (MFCC) nebo mel-spektrogramy. Tyto featury reprezentují hlasitost různých frekvencí v čase a jsou odvozené ze spektra řeči přes Fourierovu transformaci. Cílem je zachytit důležité vlastnosti zvuku podobně jako lidský sluch.

What is the function of the acoustic model in a pipeline ASR system?

Akustický model převádí zvukové vlastnosti (např. spektrogramy) na pravděpodobnosti jednotlivých fonémů nebo jejich kontextových variant. Modeluje vztah mezi akustickými daty a jazykovými jednotkami, např. fonémy. V tradičním systému to býval HMM s GMM, dnes spíše neuronová síť.

What's the function of a decoder/language model in a pipeline ASR system?

Dekodér kombinuje výstupy akustického modelu s jazykovým modelem a hledá nejpravděpodobnější sekvenci slov. Jazykový model určuje, jaká slova na sebe pravděpodobně navazují, a pomáhá vybrat správný význam při více možnostech. Dekódování se často provádí pomocí algoritmu Viterbi nebo beam search.

Describe the architecture of an end-to-end neural ASR system.

End-to-end systém modeluje přímo pravděpodobnost textu daného zvukem bez potřeby fonémového slovníku. Používá se architektura encoder-decoder s pozorností (např. LAS) nebo CTC/Transducer, která přímo převádí zvukové featury na text. Takový systém se učí z dvojic audio–transkript a často dosahuje lepší přesnosti, ale je obtížněji upravitelný.

11 Text-to-speech Synthesis

How do humans produce sounds of speech?

Lidská řeč vzniká proudem vzduchu z plic, který rozechvívá hlasivky a vytváří základní zvukový signál. Tento signál je dále tvarován rezonančními vlastnostmi vokálního traktu, jehož tvar se mění pomocí jazyka, rtů, čelisti a měkkého patra. Výsledkem jsou různé zvuky, které odpovídají jednotlivým hláskám.

What's the difference between a vowel and a consonant?

Samohlásky (vowels) jsou vytvářeny otevřeným vokálním traktem a jsou typicky znělé. Naproti tomu souhlásky (consonants) vznikají při částečném nebo úplném uzavření vokálního traktu a mohou být znělé i neznělé. Typ a místo uzávěru ovlivňuje výsledný zvuk souhlásky.

What is F0 and what are formants?

F0 je základní frekvence hlasivek, která určuje výšku hlasu. Formanty jsou rezonanční frekvence vokálního traktu a představují výrazné vrcholy v akustickém spektru řeči. Zejména první a druhý formant (F1, F2) hrají klíčovou roli v rozlišování samohlásek.

What is a spectrogram?

Spektrogram je vizuální reprezentace zvukového signálu, která ukazuje, jak se mění frekvenční složky zvuku v čase. Osa x znázorňuje čas, osa y frekvenci a intenzita je znázorněna barvou nebo jasem. Používá se k analýze řeči, protože ukazuje struktury jako formanty nebo šum.

What are main distinguishing characteristics of consonants?

Souhlásky se odlišují způsobem tvoření (např. explozivní, frikativní, nosové) a místem artikulace (např. rty, zuby, patro). Dále se dělí podle znělosti –

znělé mají vibrace hlasivek, neznělé ne. Akusticky se často vyznačují šumem a rychlými změnami ve spektrogramu.

What is a phoneme?

Foném je nejmenší zvuková jednotka v jazyce, která rozlišuje významy slov. Například změna fonému /d/ na /f/ ve slově dog“ → fog“ mění význam slova. Fonémy mohou mít různé realizace (tzv. alofony), ale význam zůstává stejný.

What are the main distinguishing characteristics of different vowel phonemes (both how they're produced and perceived)?

Různé samohlásky se liší především polohou jazyka (vysoko/nízko, vpředu/vzadu) a zaokrouhlením rtů. Akusticky jsou odlišitelné podle hodnot prvního a druhého formantu (F1 a F2). Perceptuálně je člověk vnímá jako rozdílné zvuky na základě těchto vlastností.

What are the main approaches to grapheme-to-phoneme conversion in TTS?

Používají se dvě hlavní metody: výslovnostní slovníky a pravidla. Pro jazyky s pravidelnou ortografií (např. čeština) postačí pravidla, zatímco pro nepravidelné jazyky (např. angličtina) jsou potřeba slovníky. V praxi se obvykle kombinuje obojí – pravidla jako fallback pro neznámá slova.

Describe the main idea of concatenative speech synthesis.

V konkatenativní syntéze se výslovnost skládá ze segmentů reálných nahrávek – nejčastěji diphonů. Tyto jednotky se vyberou z databáze a spojí se do výsledného signálu. Pro plynulost řeči je důležité minimalizovat rozdíly mezi sousedními jednotkami.

Describe the main ideas of statistical parametric speech synthesis.

Tato metoda využívá statistické modely, které se učí z akustických a lingvistických charakteristik korpusu. Model předpovídá akustické featury z textových vstupů a vokodér z nich generuje řeč. Výhodou je menší datová náročnost a větší flexibilita oproti konkatenativní metodě.

How can you use neural networks in speech synthesis?

Neuronové sítě se používají k predikci akustických vlastností přímo z textových vstupů. Mohou nahradit HMM i vokodéry a umožňují přímou generaci zvukových vln (např. WaveNet). Moderní přístupy jako Tacotron spojují sekvenci na sekvenci modely s generátory zvuku pro vysoce přirozený výstup.

12 Chatbots

What are the three main approaches to building chitchat/non-task-oriented open-domain chatbots?

Hlavní přístupy jsou: pravidlové (rule-based), které používají ručně psaná pravidla a klíčová slova; vyhledávací (retrieval-based), které hledají podobné předchozí věty a vrací odpovědi z databáze; a generativní (generative), které používají neuronové sítě k vytváření nových odpovědí. Každý přístup má své výhody i nevýhody, často se kombinují.

How does the Turing test work? Does it have any weaknesses?

Turingův test zkoumá, zda je člověk schopen rozeznat, zda komunikuje se strojem nebo člověkem. Pokud stroj dokáže člověka přesvědčit, že je také člověk, testem projde. Slabinou testu je, že hodnotí jen chování, ne skutečné porozumění nebo inteligenci, a je snadno zmanipulovatelný.

What are some techniques rule-based chitchat chatbots use to convince their users that they're human-like?

Pravidlové chatboty často napodobují role, kde je omezené očekávání (např. terapeut, paranoik), používají opakování a reformulaci uživatelova vstupu, nebo obecné fráze typu pokračuj, zajímá mě to". Také využívají zpětné vazby jako chápu" nebo to je zajímavé" pro udržení dojmu lidskosti.

Describe how a retrieval-based chitchat chatbot works.

Retrieval-based chatbot má uloženou databázi dialogů a při vstupu od uživatele hledá podobnou větu v korpusu. Poté vrátí odpověď, která na tuto podobnou větu navazovala. Pro zvýšení kvality se výsledky často znovu seřazují (reranking) podle relevance a kvality.

How can you use neural networks for chatbots (non-task-oriented, open-domain systems)? Does that have any problems?

Neuronové sítě mohou být použity pro generování odpovědí na základě předchozí konverzace (např. seq2seq nebo Transformer modely). Výhodou je schopnost generovat originální odpovědi, ale problémy zahrnují repetitivnost, nejasnou osobnost, halucinace a nekonzistentní odpovědi. Modely se často vrací k bezpečným a nudným odpovědím.

Describe a possible architecture of an ensemble non-task-oriented chatbot.

Ensemble chatbot kombinuje více subsystémů: pravidlový modul pro citlivé nebo časté dotazy, vyhledávací modul pro vtipy nebo fakta, a generativní modul jako záloha. Rozhodovací mechanismus vybírá odpověď podle tématu nebo skóre relevance. Často se používá sdílené NLU pro lepší pochopení vstupu.

What do you need to train a large language model?

Potřebuješ velké množství textových dat (stovky miliard tokenů), výpočetní prostředky (GPU/TPU), a dostatek času na trénování. Dále je potřeba předzpracování dat, architektura modelu (např. Transformer) a případně doladění pomocí RLHF nebo instrukčního ladění. Finetuning může být volitelný, ale je často náročný.

What are some issues you may encounter when chatting to LLMs?

LLM může znít přesvědčivě, ale často halucinuje – tedy generuje nepravdivé informace. Mívá tendenci vyhovět uživateli i při nesmyslných nebo nepravdivých požadavcích. Také má problém říci nevím" a může být náchylný ke změnám názoru nebo nekonzistenci.